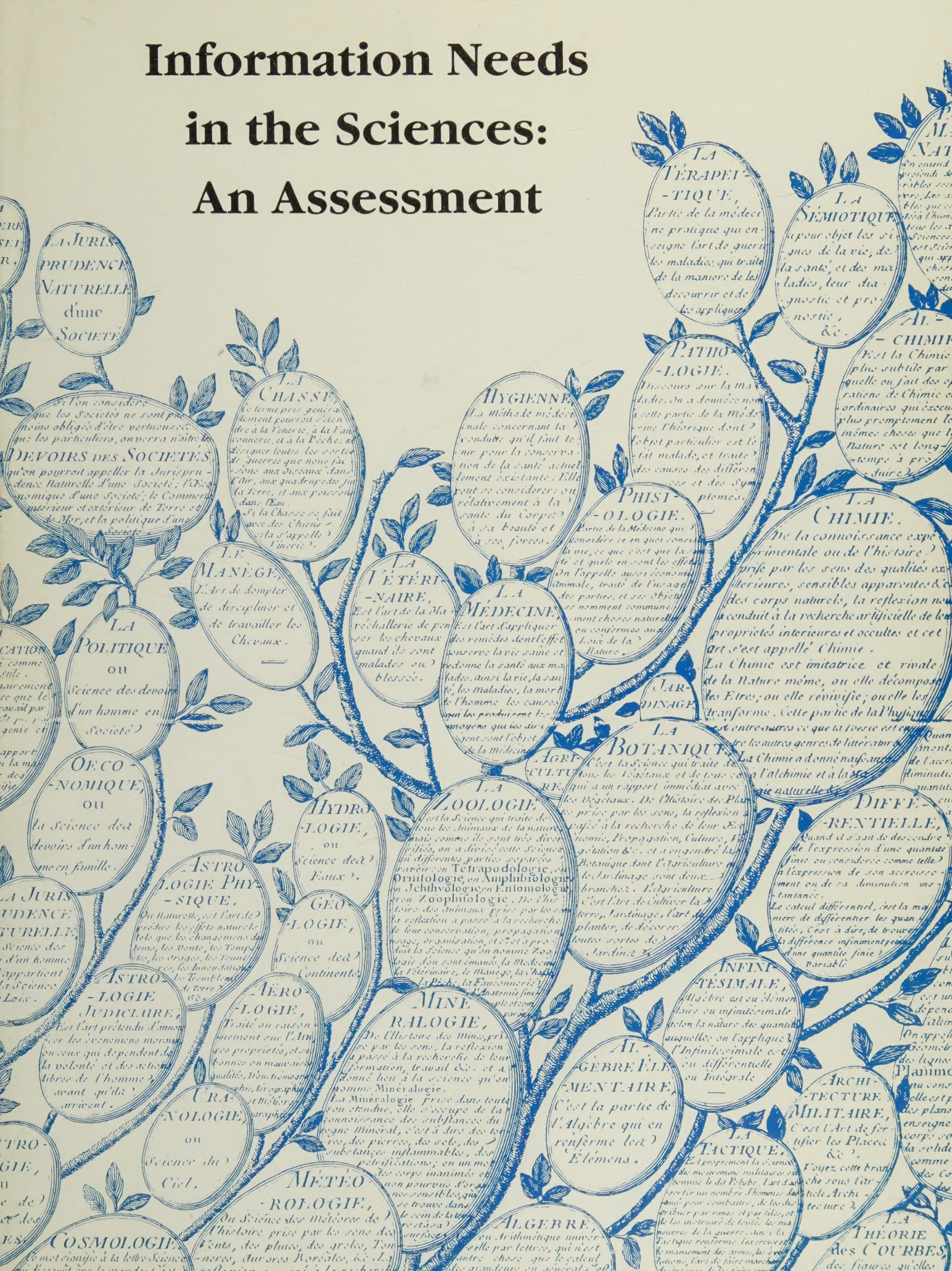
Information Needs in the Sciences: An Assessment



RLIN® is a registered service mark of
The Research Libraries Group, Inc.
Mountain View, California 94041-1100

© 1991 by The Research Libraries Group, Inc.
Mountain View, California 94041-1100
All rights reserved

First published in March 1991
Printed in the United States of America

Q
179.96
G68
1991

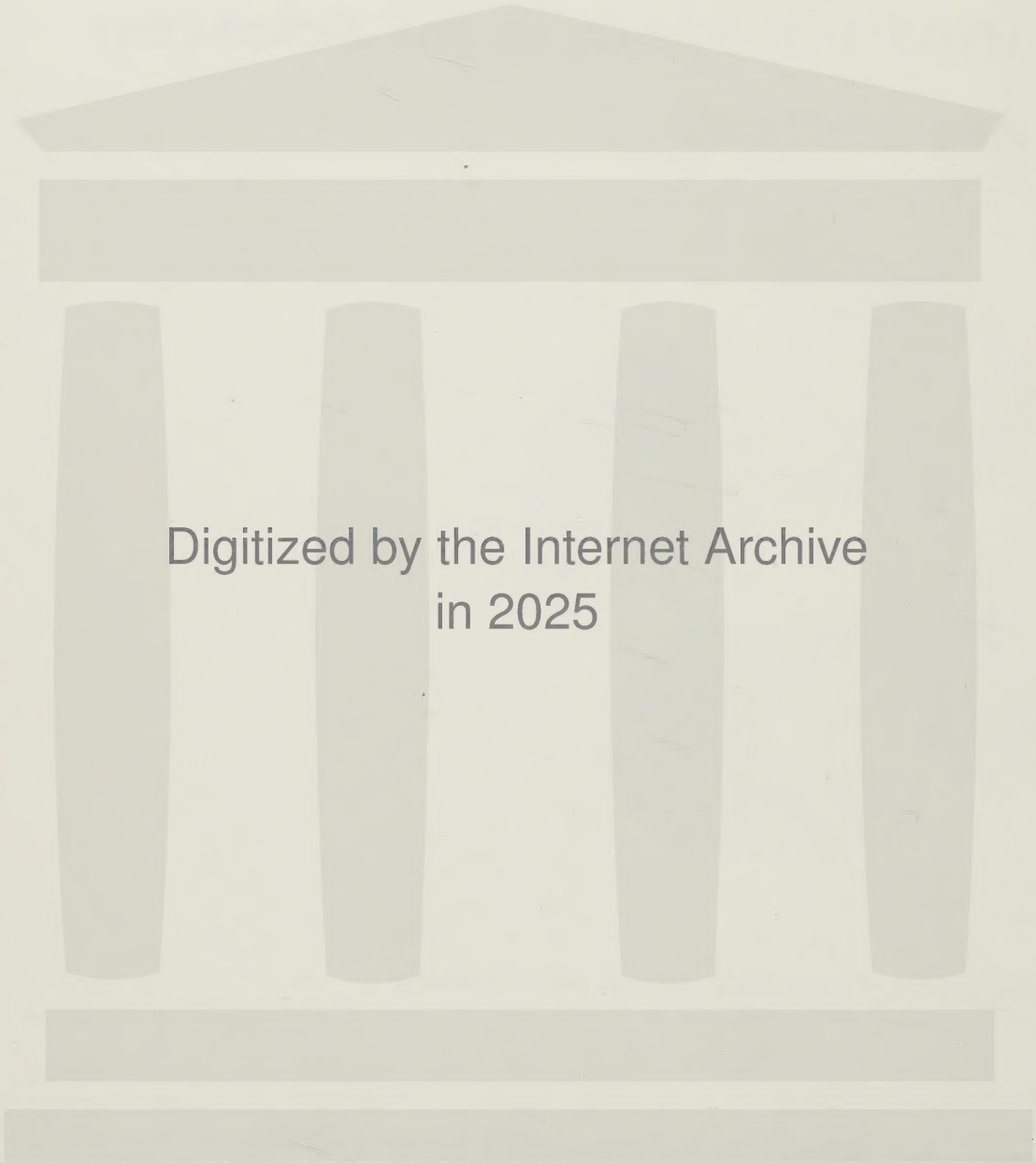
Information Needs in the Sciences: An Assessment

By Constance C. Gould and Karla Pearce

Prepared for the
Program for Research Information Management
The Research Libraries Group, Inc.
Mountain View, California, U.S.A.

MAY 30 1991

UNIVERSITY OF ARIZONA LIBRARY


Digitized by the Internet Archive
in 2025

https://archive.org/details/informationneeds0000cons

Contents

Introduction 1
Physics 3
Chemistry 12
Biology 24
Geosciences 35
Astronomy 47
Engineering 54
Mathematics 63
Computer Science 69
Conclusion 74
Appendix 77

30
6
16
17
4
5
6
7
9
10
11
12
13
14
15
19
41
43
40
32
31
50

Introduction

This assessment of information needs in eight science disciplines was produced to guide the activities of the Research Libraries Group's Program for Research Information Management (PRIMA). RLG initiated PRIMA in 1985 in response to the changing information environment at research institutions: increasing demands from faculty and students for access to machine-readable information of all kinds; and a proliferation of microcomputers making the manipulation of machine-readable information possible. PRIMA's mission is to identify the types of information researchers need better access to and to foster activities that encourage the development of new data resources.

The purpose of this assessment is to obtain a broad view of the shape of each discipline—how its dominant concerns have evolved over the last 15 to 20 years, and what its new frontiers are—and to illuminate the relationship between these trends and the data requirements of the discipline. It will also provide a basis for fostering or adopting projects regarded by researchers in one or more disciplines as valuable or unique information resources.

Interviews and consultation with 131 individuals who teach, conduct research, or are closely connected with scientific research were the basis for this document. Eighty of those consulted were faculty at 16 RLG member institutions; the remainder were primarily academic and corporate librarians. Disciplines included in the science survey are physics, chemistry, biology, geosciences, astronomy, engineering, mathematics, and computer science. A survey of the humanities was published in September 1988, and a survey of the social sciences in July 1989.

It must be emphasized that this project has produced a *preliminary* needs assessment. To discern shared elements within and among the disciplines, interviews covered a common set of topics, centering on access to broad categories of information. By its very nature, the survey could not embrace every aspect of every discipline.

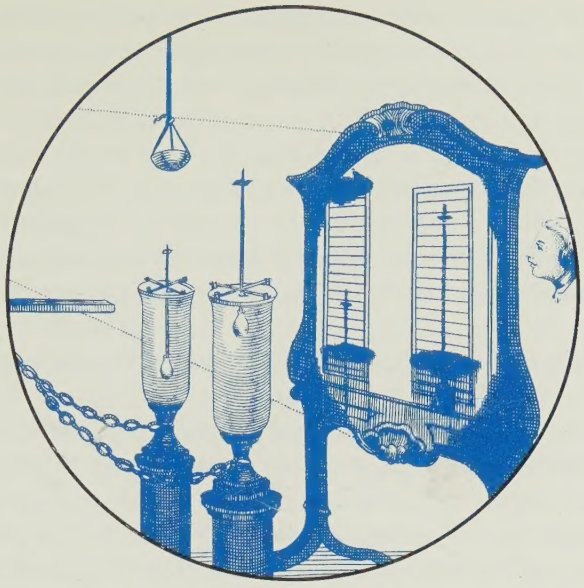
Also, many important science information resources are used by scholars in a number of disciplines. Some repetition was unavoidable, but the survey does attempt to minimize redundancy. Thus, a resource relevant to chemistry, biology, and engineering may be described fully in only one of those sections.

Additionally, although every effort was made to identify projects and resources of significance, there will inevitably be omissions. This fact is due both to the sheer number of resources in the sciences and to the frequency with which new ones appear.

Finally, the survey does not identify every information requirement but emphasizes information needs that an organization of RLG's capabilities can address.

The following summary should be regarded as a map that will assist in determining future directions for PRIMA. (The selection of projects must also take into account other criteria, such as funding opportunities and participation of prospective partners.) This summary may also be of value to organizations whose interest in improving the management of information overlaps with that of RLG; it may suggest possible projects or indicate areas for potential collaborative activity. RLG offers this assessment to the research community in the hope that it will stimulate cooperative efforts to improve access to the information vital to research.

—Constance C. Gould



Physics

General Characteristics

Nature of Research

Physics, the basic physical science, asks fundamental questions about energy, the structure of matter, and their interactions. The American Physical Society divides the discipline into the following subfields: chemical physics, condensed matter, atomic, molecular, and optical physics, fluid dynamics, particles and fields, astrophysics, plasma physics, and high-polymer physics. (In this report, astrophysics is subsumed under astronomy.)

As in the other basic sciences, the lines between research in physics and the other sciences are often indistinct. Interdisciplinary work overlapping with chemistry and biology is common, as is close interaction with fields such as materials science, electrical engineering, and geophysics.

Research in physics ranges from "big science," involving expensive facilities such as particle accelerators, large computers, and dozens or even hundreds of physicists at a number of geographically distant laboratories collaborating on one project, to projects carried out by small teams in a single laboratory. Other physicists, working outside the laboratory, aim to describe natural phenomena in mathematical form. Although large projects tend to have a high profile, the majority of physicists are not engaged in big science.

Due to the high cost of technological support for big science, researchers in these fields are heavily dependent on the research priorities of and funding from the federal government.

Nature of Information

The literature of physics is found largely in journal publications and is covered comprehensively by well-established indexing and abstracting services. With chemistry, physics has the best organized literature in the sciences.

The development of instruments that generate massive quantities of data and computers with an ever-increasing capacity to manipulate them has changed the character of research in physics, making possible data analysis and calculations that formerly would have taken years to complete.

Physics literature is shaped by the leading physical societies in the U.S., Western Europe, and Japan, whose research publications dominate the field. In the U.S., most physics literature is published either directly by the American Institute of Physics or by one of its ten member societies.

The Future for Physics Information

Physics covers a vast sweep of concepts and phenomena ranging in scale from the cosmic to the infinitesimally small. The trend to extend further the frontiers of knowledge in the field has accelerated in the past 40 years. Now, the principal limits foreseen are more economic than technical, as the equipment and instrumentation developed in the pursuit of this goal have become increasingly sophisticated, complex, and expensive. With research facilities like the superconducting supercollider reaching cost levels measured in billions of dollars, big science has come under attack. Thus, it is possible that support for big science may level off or even decline in the near future, bringing with it a concomitant moderation in the generation of new physics information in those fields.

In the past, areas of research originally in the domain of physics have been taken over by other academic fields. For example, acoustics is now largely subsumed by engineering, as is semiconductors, once the province of condensed matter physics. In the future, theories and data developed by physicists will continue to become useful to other academic fields and to be applied by industry.

Use of Physics Information

A notable characteristic of physics is the sharp distinction between theoreticians and experimentalists. The gap between these two groups is probably more pronounced than in any other scientific discipline and is reflected in the kinds of information sought by the two groups. Theoreticians rely most heavily on the work of other theoretical physicists, that is, their predecessors who have worked on any given problem of interest. The experimentalists are more concerned with questions involving the way in which experimental procedures are carried out, the measurements recorded, the accuracy and precision of the measurements.

Information useful in answering theoretical and experimental questions can be located in the standard indexing and abstracting tools, in either the print version or online. In the case of questions concerning physical properties of materials, the standard reference sources are also useful. However, where research ventures into new territory, such as the mechanical or thermodynamic properties of materials at very low temperatures, published data may be sparse.

Primary Literature

Serial Literature

Journals remain the principal vehicle for the dissemination of physics research information. As in all science disciplines, physics has experienced an increase in the number of articles published and in the size and cost of journals. Physics librarians report that the average price of physics journals is high—although not as high as that of chemistry journals—and accounts for an increasing proportion of the acquisitions budget.

Despite the international character of physics, the English language predominates in scholarly publications. English translation journals published by the American Institute of Physics play an important role in making major Russian, Chinese, and Japanese research available.

Monographs

Monographs are indispensable as graduate texts, as reference tools for research or laboratory equipment design, as an introduction to new research areas, and for review and consolidation of ideas and developments in specific subject areas. But in comparison to the humanities, this body of literature is relatively small, since fewer books are published in the sciences and only a minority withstand the test of time to become classics.

An important segment of monographic literature—books published after the mid-19th century—is deteriorating rapidly because it is printed on high-acid paper. Several research libraries have undertaken to microfilm volumes in the history of science, including physics. However, because of the widespread misconception that research scientists do not need older materials, preservation microfilming of these books in general receives a low priority.

In addition to monographs, published conference proceedings are very important in physics. Proceedings that in earlier days would have been produced inexpensively by mimeograph or photocopy are now formally published and sold at a correspondingly higher price.

Major Indexing and Abstracting Services

Four comprehensive indexing and abstracting services attempt to cover the entire body of physics literature. The two most important are *Physics Abstracts*, produced by the Institution of Electrical Engineers, and *Physics Briefs*, produced by the Fachinformationszentrum Energie, Physik, Mathematik (commonly called FIZ Karlsruhe) in cooperation with the American Institute of Physics and Deutsche Physikalische Gesellschaft. The other two are *Bulletin Signalétique*, produced by the Centre National de la Recherche Scientifique in Paris, France, and *Referativnyi Zhurnal Fizika*, produced by the Academiia Nauk in the USSR.

Print Indexes

Science Abstracts, Section A, *Physics*, or *Physics Abstracts*, in publication since 1898, covers the full range of physics literature published in all languages of the world. Abstracts are arranged in ten subject areas, with both author and subject access provided. Although coverage emphasizes journal articles and conference papers, selected books, technical reports, and dissertations are also included. A thesaurus aids information retrieval. *Physics Abstracts* adds entries at a rate of approximately 149,000 a year.

Physics Briefs covers the worldwide literature in all fields of physics and related topics, and was first published in 1845 under a different name. As with *Physics Abstracts*, arrangement is by subject, with both author and subject access provided. Special emphasis is given to nonconventional literature and to publications in Eastern European languages. Items are added at a rate of about 120,000 a year.

Many physicists also use *Science Citation Index* (*SCI*), which attempts to cover the breadth of science and technology literature, including the core physics journals. In all, some 2,500 journals are included, which is not comprehensive coverage but still useful. Published by the Institute for Scientific Information, *SCI* is valuable for its system of recording who cites whom in any given field of science. This information allows a researcher to track the development of specific topics by different scientists over the years. It is also useful for interdisciplinary work.

Electronic Indexes

More than 3.3 million citations with abstracts dating from 1968 are accessible in the online database INSPEC, available through BRS, DIALOG, STN, and a number of other vendors. INSPEC includes not only *Physics Abstracts*, but *Electrical and Electronics Abstracts* and *Computer and Control Abstracts* as well. Over 4,000 journals are covered, plus conference proceedings, books, reports, and other types of publications.

The online version of *Physics Briefs* is available through STN. It contains some 1.3 million citations with abstracts to literature published since 1979. Although it covers a smaller number of journals than INSPEC (approximately 2,800 titles), some users find it more timely and attribute this feature to the involvement of the American Institute of Physics, which publishes a number of the journals indexed. The searching capabilities available through STN also add to its usefulness.

Science Citation Index is available online as SCISEARCH and on CD-ROM.

Use of these databases is to a certain extent dependent on their accessibility and cost; advantageous pricing arrangements such as flat fees for use can result in a significant increase in use. Yet many physicists bypass these resources altogether and head for the serials shelves of the library, where they browse through the latest copies of their favorite journals, both to stay current in their own area of research and to keep up with trends in the discipline as a whole.

One complaint about online bibliographic searches is that, because databases are so big, even a well-defined search may yield an unmanageably large result. Information scientists have acknowledged this problem ever since Boolean searching became the norm for online retrieval. At the other extreme, however, the ease of retrieval of an obscure reference that might have required hours of searching to find in print indexes helps to balance the scales.

Current Awareness

The information most important to physicists is often too recent to have been published, particularly in the subfields of nuclear and particle physics. The recent explosion of research in the field of cold fusion illustrates clearly the inadequacy of established means of scholarly communication in a field where results are being produced weekly or even daily. In these instances, results, both real and questionable, pour out so fast that journals attempting to keep up are out of date before they are printed. Only instant means of communication will work: transmission via computer networks or fax, or at a gathering of scientists. In these cases, communication depends heavily on informal networks since there is no central source of information.

Electronic Networks

Probably the most important and heavily used current awareness resources for physicists are electronic networks such as BITNET. These networks allow physicists to keep in constant communication with each other worldwide, reducing the time this takes from weeks to days or even hours. Physicists appear to be among the most active and enthusiastic users of electronic networks. The explosive growth in this type of interchange implies a more casual approach to the acquisition of information and may ultimately pose a threat to the orderly reporting and maintenance of the records of research results.

In most cases, research proceeds at a less frantic pace. For more usual circumstances, several expedients have been developed.

Letters Journals

Whole journals or parts of journals devoted to letters were designed for rapid communication of research results. These letters are contributed by the researchers themselves, are not refereed, and describe current research or preliminary results. But many physicists complain that letters journals have also begun to suffer from a long time lag between submission of results and appearance in print.

Newsletters

Newsletters are a less formal and more timely source of information on research in progress. The Ames Laboratory's *High Tc Update*, created to provide quick access to information on research in superconductivity, is a well known example. It includes descriptions of projects and results of experiments and lists preprints and contact information.

Technical Reports

Technical reports, which contain the results of government-sponsored research, are used in some subfields of physics. Groups issuing technical reports include national and international laboratories such as Fermilab, CERN (the European Laboratory for Particle Physics), Stanford Linear Accelerator Center, Brookhaven National Laboratory, and Sandia National Laboratories, where important work is carried out. Usually not written for publication, their technical reports often contain more data than preprints intended for a journal.

The National Technical Information Service (NTIS) has been designated as the central collection point for all reports describing federally sponsored

research as well as research sponsored by foreign governments. However, researchers say that NTIS misses important foreign reports and laboratory series describing research not directly sponsored by the government. Some suspect that sponsored researchers outside the federal laboratory system also neglect to send their reports to NTIS.

Many libraries do not collect technical reports systematically and may not retain older reports, although some libraries have established extensive collections of technical reports in specific areas, generally outside the NTIS collections. Unfortunately, information about individual library collections of technical reports is hard to come by.

Preprints

Preprints, in effect prepublication reports about ongoing work, are a vital means of communication. In the interest of rapid dissemination, preprints are not subjected to the rigors of peer review and therefore do not have the same status as a paper in one of the refereed journals. Their value to the authors is that they establish priority of discovery; they also accelerate the pace of research.

Most are published in English, either originally or in translation. The volume of preprint output can be high; some physicists report receiving as many as 10 or 12 a week. The volume tends to vary depending on the pace of developments. Because preprints are not available through the normal channels and are generally considered to be ephemeral information, most physics libraries do not collect preprints comprehensively although they may maintain collections in specific subject areas. Researchers usually obtain them through exchanges between academic departments or informal arrangements of their own, but these methods are unreliable and inefficient.

Because experiments in high-energy physics are so expensive, physicists can't wait for formal publication to get information about related work, so this group in particular has long relied on preprints to communicate experimental results and speculation about experiments.

Exchange of information about preprints in high-energy physics is facilitated by a weekly newsletter published since 1969 by the Stanford Linear Accelerator Center (SLAC) Library called *Preprints in Particles and Fields*. Kept simple for the sake of speed of publication, the newsletter lists the author, title, and local institution's report number of some 150 preprints each week. It also lists *anti-preprints*, published articles that first appeared as preprints. This information is also available on BITNET or the Internet (another communications network linking major research institutions) from the HEP (High Energy Physics) database at Stanford. Documents themselves must be obtained from the author.

Preprints are valued because they provide almost instantaneous publication of research. However, this excessive haste to disclose results can lead to the inclusion of errors or incompletely formulated ideas that may be misleading or totally erroneous. This problem has not figured prominently in the high-energy field, but it has caused confusion in others like

cold fusion, where some research results have turned out to have little scientific validity. However, no solution has yet been found for reconciling the requirement for speed with that of reliability.

Databases

The Superconductivity Information System (SIS), established and maintained by the Department of Energy's (DOE) Office of Scientific and Technical Information, attempts to answer the need for rapid access to a centralized source of information for researchers in superconductivity. Available through a dial-up connection, SIS offers three databases. The most heavily used are the Preprints Data Base and Energy Data Base. The first contains abstracts of papers on superconductivity submitted to journals or written for presentation at a conference. Although the database provides a facility for direct input of information by researchers, most records are obtained from sources such as the American Institute of Physics and the Ames Laboratory, a center for superconductivity research. Since the emphasis is on rapid dissemination of information, papers have not been refereed. Some researchers check the file as often as several times a month.

The Energy Data Base component of SIS is actually a subset of DOE Energy, a bibliographic database containing some 2.5 million citations to literature in all fields of energy published since 1974 (plus selected older literature) and also available through STN and DIALOG. Coverage is worldwide, with an estimated 55% of entries coming from foreign sources, and includes technical reports, journal articles, monographs, conference proceedings and papers, and patents. The portion offered by SIS includes information added to the database in the most recent 14 months.

Other portions of SIS, used less heavily than the two above, include Work in Progress, the core of which is current information on DOE research contracts in superconductivity. As in the case of the Preprints Data Base, researchers may enter information themselves. Finally, SIS provides an online bulletin board and electronic mail service to communicate news from DOE and other federal agencies, information on meetings and conferences, announcements of work in progress, and information from professional societies.

Resources such as Stanford's HEP database and SIS provide an important service to researchers: timely access to a centralized source of information about current research in one subject area and related information and communication facilities.

PI-NET, a service of the American Institute of Physics, provides information about papers accepted for publication months before they actually appear.

Current Contents, a print collation of tables of contents, fully indexed, from the journals that are indexed in *SCI*, is popular with some physicists because it provides quick access to information about the most recent published research. It is also available online as Current Contents Search.

Handbooks and Compilations

Compared to researchers in other fields such as chemistry and engineering, physicists are not heavy users of reference information. Often the information they seek is likely to be so specialized and technical that it is unlikely to be found in a reference source.

However, standard reference tools are sometimes useful, particularly for information on physical properties of materials. These tools include the *JANAF* (Joint Army, Navy, Air Force) *Tables of Thermochemical Information*; the *National Standard Reference Data Series* of physical and thermochemical property information and its predecessor, the *International Critical Tables*, published by the AIP, the Chemical Society, and the National Institute of Standards and Technology (formerly the National Bureau of Standards); as well as tables of infrared and ultraviolet spectra. The combined 6th edition and "New" series of the *Landolt Bornstein: Numerical Data and Functional Relationships* is an immense compilation of physical, physico-chemical, and astronomical data that currently consists of 168 volumes and is still growing. All of these tools aid in identifying materials and understanding their interactions.

The *JANAF Tables of Thermochemical Information* is available online through STN as the Joint Army-Navy-Air Force file, or JANAF. It provides a compilation of critically evaluated chemical thermodynamic properties of inorganic substances. An important group of specialized online databases dealing with nuclear structure and reactions is maintained by the National Data Documentation Center (NDDC) at Brookhaven National Laboratory.

Other Electronic Resources

Physicists' research information needs go far beyond the bibliographic, and involve the storage, analysis, and transmission of primary research information. Data produced by high-energy accelerators provide the most notable example. Accelerators are located in only seven laboratories in the U.S., Europe, the Soviet Union, and Japan, but physicists who use them—often in cooperative projects—are scattered among universities and research institutions around the world.

Networks allow for remote data collection, sharing of information among colleagues, effective collaboration on complex experiments, and the production of co-authored papers. Good networking facilities also reduce the need to travel to accelerator sites. Advances like the development of the proposed superconducting supercollider will result in greatly increased demands for networking and data analysis. Even now, data from only a minute fraction of the millions of particle collision events a year produced by accelerators can be stored and analyzed.

Future Needs and Directions

Physics Literature

Easy and quick access to published literature, especially recent serials, will continue to be a key requirement. The major indexing and abstracting services provide excellent access to information about published literature.

Unpublished literature will continue to be essential. Cooperative projects among libraries that collect technical reports would improve access to those not included in NTIS. These efforts should be coupled with an effi-

cient and inexpensive document delivery capability. More reliable means of obtaining preprints, especially in fields outside high-energy physics, are called for. Again, cooperative projects among libraries that collect preprints, tied to a document delivery capability, could meet this need. Some physics librarians suggest that electronic distribution of the texts themselves on computer networks would be useful.

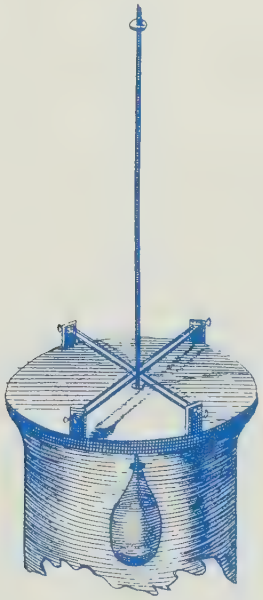
Electronic Resources

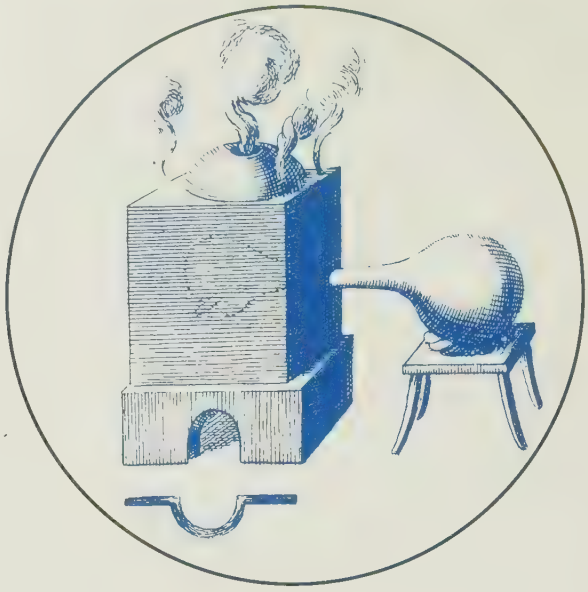
The importance of networks will continue to grow. Although high-energy physicists generally have access to good networking facilities, their demands are likely to increase. The needs of physicists in other subfields, who often do not have access to the same level of networking facilities as their colleagues, are also likely to increase.

Current awareness, a critical need in many fields of physics, can be supported by networks. Particularly because physics research tends to remain in relatively well-defined subject areas, opportunity may lie with centralized sources of information about a wide array of resources in one subject area. These could include some or all of the following: abstracts and references to recent publications, preprints, and technical reports; information on conferences and meetings; online document-transmission, electronic-mail, and bulletin-board facilities; information from professional societies; and information on how to obtain access to related databases.

Preservation

Increased attention should be given to the preservation of deteriorating physics monographs.





Chemistry

General Characteristics

Nature of Research

Chemists investigate the properties, composition, and structure of matter and its transformation from one kind of substance into another. Research in this discipline can be divided into the subfields of organic, inorganic, physical, analytical, theoretical, nuclear, biophysical, biological, medicinal, environmental, and macromolecular chemistry; however, these overlap a great deal.

Research in chemistry is far removed from the "big science" model. Chemists generally work in small teams and are involved with every aspect of an investigation. Publications in the field rarely involve the large number of authors common in some science fields. Cross-fertilization between the industrial and academic sectors is frequent.

Research in chemistry is highly dependent on what is already known and reported in the literature. Consequently, research chemists are heavy users both of libraries and online information sources. Chemistry libraries are often located near research laboratories, allowing for quick consultation of reference sources while an experiment is running.

Nature of Information

Chemical information is distinctive in that it deals with atomic and molecular species, which are precisely and unambiguously defined by their molecular structure. Since chemical and physical properties do not change over time, older literature is as essential as current literature.

The literature of chemistry is accessible through numerous access points. In addition to subject and author, both print and electronic indexes offer searching by chemical names, molecular formulas, and patent numbers.

Online indexes offering the further advantage of searching by molecular structure, molecular substructure, or structural transformation (chemical reaction) have made the vast chemistry literature much more accessible and manageable than it would otherwise be. Hence, despite the great diversity of research in chemistry and the enormous body of literature it has generated, chemistry literature is relatively well defined and controlled.

Chemistry has the advantage of some of the most advanced information resources and technology now available, and computers have already become common tools for both laboratory work and literature searching. In addition to bibliographic files, chemistry is now served by full-text files, numeric data files, and chemical reaction files.

The Future for Chemical Information

Chemistry is the basis of a large national and international industry that invests heavily in basic research. To remain competitive, the chemical industry must have access to the latest technology and information.

Chemical information is of fundamental importance to other scientific disciplines. As a result of advances in chemistry, the nature of research in biochemistry and genetics has changed radically, and medicine is increasingly explained in terms of molecular structure. Both environmental and materials research depend on chemistry.

Resources of continually increasing magnitude will become available to the chemical information industry. Measured by the number of known chemical compounds, the growth in chemical knowledge has been exponential in the last 20 years and shows no signs of abating.

Trends also indicate that chemists and researchers from related disciplines will make ever-greater use of electronic information sources, resulting in a concomitant decline in dependence on printed indexing and abstracting tools.

Use of Chemical Information

Chemists and scientists in cognate fields use chemical information in three main ways: (1) to answer a specific question, (2) to gain background knowledge in a narrow area of interest (in-depth searching), and (3) to keep abreast of key trends in wider fields, thus gaining ideas that will apply to their own work (browsing).

Specific questions that chemists seek to answer can be divided into types. Some of the more representative are: Has a specific compound been reported? Have any examples of a specific class of compound been reported? Is a particular property of a specific compound (e.g., vapor pressure at a given temperature) known? How can a property be best estimated by analogies from the literature? Has a particular type of reaction ever been reported? What synthesis methods have been reported for a specific compound? Can a compound with certain defined properties be synthesized? To which compound does this spectrum (or other set of properties) refer? What patents describe a given process?

The chemistry literature containing the answers to these questions is vast. More than 10 million compounds have been synthesized, and the variety of physical properties recorded is also huge. Even 50 years ago, searching by formulae index was possible and was far more effective than searching by name, author, or title. However, as the number of known compounds grew this type of searching was no longer practical.

The advent in the 1970s of online services that offer searching by chemical structure has made it possible for chemists to make use of the burgeoning quantity of chemical information. In fact, this capability has become essential to substantive research in chemistry.

Chemists can now, on the computer screen, draw or type in a two-dimensional representation of a molecule or part of a molecule and then search a database for information about that molecular structure or substructure. This is a powerful way to retrieve information about chemical substances, particularly families of substances that share structural features. Moreover, structural searches can be made without recourse to complex chemical nomenclature or molecular formulas.

Thus, a chemist would determine whether a specific compound has been reported by conducting a full structure search for the chemical structure that is unique to the compound. Examples of a specific class or family of compound are identified by substructure searching.

Queries about properties, by contrast, are less easily answered. They can be addressed in two ways: the first is via a comprehensive structure or substructure search followed by a manual search of each record for the required information. The disadvantage is that the number of records could be very large, making the manual search very time-consuming. These questions can also be approached by using reference aids, none of which are comprehensive, leaving open the possibility that information may be missed. The increasing availability of online numeric files should greatly enhance the search for property data.

In-depth searches can be carried out with two main strategies, either by answering a series of specific questions of the type mentioned above or by the more traditional method of looking for tertiary literature (reviews and the like) in the general field. However, this category of literature is usually years behind the primary literature.

Primary Literature

Serial Literature

Serial literature is the most important medium of communication in the discipline and the ultimate source of chemical and physical property information. Fast access to current literature certainly ranks first among chemists' information needs, but access to retrospective literature is probably a close second. The discipline's dependence on serial literature has notable economic consequences. Chemistry serials have one of the highest average costs per volume of science disciplines; in fact, between 80% and 90% of a chemistry library's materials budget is typically allocated to serials.

Access to the primary literature is now overwhelmingly obtained through the print copy. Online access to the full texts of some serials in chemistry is available from commercial services, but, although bibliographic databases have gained wide acceptance in the discipline, full-text databases have so far had little success. Chemists cite several reasons for their lack of enthusiasm: the journals offered online are also those most commonly available in print; the cost of online access is high; finally, and most significantly, online journals do not include diagrams and illustrations and some of the most important information is therefore lost. The greatest benefit of full-text databases is the searching capability they provide, but at the prices charged for the databases, this is a luxury few can afford. Finally, online journals are unlikely to replace hard copy because chemists like to browse actual issues of journals, either to search for particular molecular structures or to see who is publishing what.

Projects currently in progress may diminish resistance to electronic distribution of information. At least one journal is issued exclusively on diskette. Another project, begun in 1990 at Cornell University, will attempt to improve on present online distribution by making a seven-year cumulation of full texts, including formulae and graphics, of 20 American Chemical Society journals, available on PCs throughout the chemistry department. The project will determine whether chemists actually use the system and what searching software functions best.

Patent Literature

Patents are vitally important to research in chemistry. They are often the earliest source of chemical information and sometimes the only source, particularly for certain data such as illustrations and experimental details. Patents are also an important source of ideas, allowing chemists to learn about new processes or uses and to get a glimpse at what chemists in other laboratories are doing. Corporate research chemists use patent literature frequently.

For several reasons, academic libraries devote less resources than do corporate libraries to providing access to this literature. The patent literature is extremely complex, requiring specialized training and frequent use to maintain facility in searching it. Patent databases are among the most expensive of information resources. Finally, there is less call for patent literature in academic libraries, since academic chemists make their research widely available and seldom seek patents.

Two categories of patent literature exist: quick-issue patents, which are unevaluated patent applications issued in some countries to ameliorate the problem of backlogs in the granting of patents, and final patents, which have undergone thorough review. Patent systems are complex, varying from one country to another, and the volume is enormous: thousands of patents are published each week worldwide. As a consequence, copies of patents are generally more difficult to obtain than serial literature, with expert assistance often necessary to locate them.

Major Indexing and Abstracting Services

A comprehensive abstracting service attempts to abstract the whole of the chemical literature. By far, the most important is *Chemical Abstracts*, produced by Chemical Abstracts Service, a division of the American Chemical Society. British and German abstracting services ceased in 1953 and 1969 respectively. The only other major service still in operation is a Russian publication, *Referativnyi Zhurnal, Khimiya*, sometimes referred to as *Viniti*.

An abstracting service is used in two related ways: as a source of information (contained in the abstract) and as a way into the primary literature (journals and patents) through the reference.

Print Indexes

Chemical Abstracts (*CA*), produced by the Chemical Abstracts Service (CAS), provides the principal access route to chemistry literature published since 1907, when the index began publication. It contains descriptive abstracts in English and now covers articles from some 14,000 journals, monographs, conferences, technical reports, and patents from 29 countries. Its worldwide service collects information from more than 150 nations, adding some half-million chemical compounds a year.

The *CA* print indexes provide access by author(s), general subject, chemical substances by name, molecular formulae, ring systems, and patents. Despite the success of an online version, the print version is still used, in part because of the expense and difficulty of learning structure searching online. Scientists in many other disciplines, for example, physicists, biologists, engineers, and physicians, also consult *CA*.

Chemists also find *Science Citation Index* useful because of its unique system of citation indexing and coverage of core journals in the discipline. (See **Physics** and **Biology**.)

The German abstracting service *Chemisches Zentralblatt*, which began publication in 1830 and ceased in 1969, is still very useful for historical coverage. Another major service, *British Abstracts*, covers the years 1926–1953.

Referativnyi Zhurnal, Khimiya, which first appeared in 1953, is useful particularly for Russian and Eastern European literature. Its importance could increase dramatically when a recently announced project to publish an online version in English comes to fruition.

The leading provider of patent information is Derwent Publications, Ltd, which produces *Chemical Patents Index*, a print publication providing worldwide coverage of chemical patents issued since 1951. Derwent groups patents into families to facilitate searching.

Electronic Indexes

CAS Online, produced by the Chemical Abstracts Service, has become the sine qua non of research in chemistry because it provides access to bibliographic information by chemical structure. It offers five files with chemical information: CA, CASREACT, CApreviews, CAOLD, and the Registry File.

CA, which corresponds to the print index *Chemical Abstracts*, includes bibliographic information on some 9 million documents published since 1967 and is available from a number of vendors (the name and chronological coverage vary with the vendor). STN offers a version of CA that includes abstracts in addition to the basic bibliographic information provided by other vendors and other information such as analytical procedures and property data.

Expert knowledge is required to use CA successfully and efficiently; many organizations employ specialists to perform all searching because the cost is so high. Academic chemists often have their own CA account and graduate student assistants who are trained to use it. With the spread of substructure searching, allowing users to describe parts of a molecule specifically and the remaining part with a "wild card," retrieval possibilities are becoming much more powerful but correspondingly more complex—and expensive.

CASREACT contains information on reactions of organic substances, including organometallics and biomolecules, selected from papers reported in 104 core journals. CApreviews is a current-awareness database with information on records that appear six to eight weeks later, with an abstract, in CA. CAOLD contains records for CA references prior to 1967.

The CAS Online Registry File is a chemical structure and dictionary database designed for substance searching and identification. It contains more than 10 million unique substance records, with each substance identified by the CAS Registry System of numbers. Records also contain the structure diagram, which can be both searched and displayed, and the 10 most recent references since 1967 citing the substance.

Patent information. Derwent provides online access to patent information through WPI (World Patents Index), which contains about 4 million citations with abstracts to chemical, electrical, mechanical, and general patents issued by 30 major patent-issuing authorities. Chronological coverage varies according to the type of chemical information; coverage of pharmaceuticals extends back to 1963, with comprehensive coverage of all chemicals beginning in 1970. Derwent also offers a patent copy service which furnishes copies of virtually any patent whether listed in the database or not.

CAS Online's CA file includes information on patents as well as serials and other types of documents. Its coverage of patents, although less extensive than Derwent's, includes a significant number of patents issued by 29 patent offices, including those of West Germany and Japan as well as the U.S. A document delivery service provides copies of recent patents.

Another major provider of patent information is IFI/Plenum Data Company, which covers primarily U.S. patents issued since 1950. Its online database Claims/Uniterm contains more than 800,000 citations to all U.S. chemical and related patents.

Two other important online sources of patent information include JAPIO (Japanese Patent Abstracts in English), which contains more than 2 million unexamined Japanese patents issued since 1976 and is said to provide information faster than Derwent. The second is INPADOC, with citations to all types of patents issued by 55 countries, the European Patent Office, and the World Intellectual Property Organization. Although coverage varies by country, the earliest patents date back to 1968.

Finally, the U.S. Patent Office produces a commercial CD-ROM, called Cassis, which lists nearly 5 million records of U.S. patents, including all patents issued since 1790. The 1.5 million patents issued since 1969 are indexed by the patent holder's name and the descriptive title of the invention; many are also accompanied by a synopsis of the technical content of the patent. Updated disks are issued every two months.

Current Awareness

It is almost a truism to observe that chemists require fast access to recent literature. Chemists have long been information conscious, often allocating time each week in the library to browse through the latest journals. Because they review grant proposals and articles, most active researchers are aware of recent research in their area of interest well in advance of publication. Informal networks are another important source of information about current literature. When the need arises, an informal network may become more formalized, as in the case of the *Quantum Chemistry Program Exchange Bulletin*, which provides a forum for the subfield of quantum chemistry. It includes brief communications on topics which do not warrant a full journal article and information on conferences and workshops, hardware and software developments, and new books of interest.

Preprints are seldom used in chemistry. Most chemists prefer to wait for an article that has been reviewed by peers. Once review has taken place, publication in some journals occurs within several months. Indexing in CApreviews, available through STN, appears six to eight weeks before the full citation and abstract have been completed for *Chemical Abstracts* and its online version, CA.

CA Selects—a print, biweekly subset of the latest articles to appear in *CA*—is used by chemists as a current information source in specific subfields of chemistry. Subscriptions are available in a wide choice of *CA* subject sections, e.g., chemical hazards and liquid crystals.

Handbooks and Compilations

Handbooks and compilations of physical data are a basic reference source for the working chemist and are heavily used. They are employed for many purposes, e.g., to identify a compound (either through its physical properties or a graphic representation of its unique spectrum), to locate patent information, or to identify preparative methods and reactions. The level of detail in chemical reference sources varies. Chemical encyclopedias and dictionaries, which cover the most important aspects of a subfield but do not attempt to be comprehensive, are useful for quick reference. Larger compilations yield greater detail and are correspond-

ingly more complex. The chief virtue of these reference sources is that they present a large body of evaluated data covering a long chronological period.

No handbook is entirely comprehensive, and thus the search for information is usually a sequential process in which a number of sources are consulted. Multiple sources are often used even if information sought is located early in the search to ensure as much accuracy and completeness as possible.

Frequently Used Handbooks and Compilations

The *Beilstein Handbook of Organic Chemistry* (*Beilstein Handbuch der Organischen Chemie*) is a multivolume reference work containing information and data related to the structure, preparation, and properties of organic compounds reported in the primary literature since 1830. It is valued for combining critical evaluation of the primary literature with an ordered presentation of verified data and facts. Originally published in German, it began to appear in English in 1984. Within each of four series (which in turn has a number of multivolume supplements) entries are ordered according to molecular structure using the Beilstein System of Compound Classification. Even so, determining which volume to consult to find a particular molecular structure can be difficult for users not thoroughly familiar with the reference work.

Major portions of *Beilstein* are now available online, enabling chemists to search for physical property and chemical structure information at the same time. Beilstein Online now covers 1.7 million heterocyclic and acyclic substances out of a total of 3.5 million substances represented in the print version. The entire corpus is projected to be available online in 1991; coverage now extends from 1830 to 1960. In addition, material abstracted from current primary literature is incorporated as it is received. As the file becomes more complete, its usefulness to chemists, physicists, and engineers will increase significantly.

In Beilstein Online, a structural file allows searching by structures and substructures. A factual file contains up to 400 items of information for each compound, including physical properties, molecular formula, literature citations, and the Beilstein Registry Number.

Chemists say that although *Beilstein* is an essential reference tool, both the print and online versions have failed to keep up with the most current literature, which is a serious deficiency.

The *Gmelin Handbook of Inorganic and Organometallic Chemistry* (*Gmelin Handbuch der Anorganischen Chemie*) is a multivolume reference work containing systematic classifications of critically evaluated data covering the fields of inorganic, organometallic, and physical chemistry from the mid-18th century. Like *Beilstein*, *Gmelin* is an open-ended publication in a number of editions, to which supplementary volumes are continually being added. Also like *Beilstein*, *Gmelin* is now published in English, with earlier volumes in German. Subject material is arranged primarily in terms of the chemical elements and their compounds.

The Gmelin Formula Index, or GFI, is an online database that indexes the handbook. It contains two components of the Gmelin handbook: the "Gmelin Formula Index" and the "Complete Catalog." Searching is by molecular formula only.

The *Heilbron Dictionary of Organic Compounds* serves as a basic reference source—literally, a dictionary—for, according to *Heilbron*, "the more interesting and important" organic substances. Entries include the name of the compound, alternative names, CAS Registry number (in the most recent edition), details of the chemical structure and physical, chemical, and other properties, and leading literature references. *Heilbron* is also available online.

The *CRC Handbook of Data on Organic Compounds* is a two-volume reference source giving the basic properties (e.g., melting point, density) of a selected number of organic compounds with references to sources of more detailed information. The *CRC Handbook of Chemistry and Physics* is also useful.

The *Merck Index* is a one-volume encyclopedia of chemicals, drugs, and biologicals, which appeared in its tenth edition in 1983. Although its original purpose was to cover organic and inorganic chemicals and drugs, the abundance of important new compounds has led the *Index* to emphasize compounds that reflect the interdependence of chemistry, biology, and medicine. It is also a good source of toxic and hazard information. Information is presented in some 10,000 short entries describing approximately 30,000 substances and providing references to the literature on them. The *Merck Index* is available online through several vendors as well as in print.

The *Wiley/NBS Registry of Mass Spectral Data* is a 1989 seven-volume publication that correlates mass spectral data with a particular compound and thus allows identifications to be made on the basis of mass spectral measurements. It attempts to be as comprehensive as possible and contains more than 110,000 compounds. The online version corresponds to the print product, augmented by new data as they become available. It also offers two interactive programs for retrieving and analyzing spectral data. The *Wiley/NBS Registry* is also available on CD-ROM.

Other Electronic Resources

Reaction Searching Databases and Software

The products in this section are not available in print versions.

CASREACT offers information on more than 700,000 single-step chemical reactions extracted from 100 key synthetic organic journals published since 1985. It includes a graphic representation of each reaction, abstracts and bibliographic references to the journal source of the reaction, the molecular formula, and up to 50 synonyms for a compound. CASREACT does not provide for true reaction searching, but is confined to structure searching on reactants and products—a significant limitation.

REACCS, a product of Molecular Design Limited (MDL), is a software package that allows the user to search for specific types of chemical reac-

tions using molecular structures and substructures. It can be used with proprietary or commercially available databases such as MDL's Current Literature File, which offers selected new organic reactions published in the recent chemical literature. REACCS and related databases are best used to explore a broad range of synthetic methods for approaching a problem—"brainstorming"—and they are highly valued for this application. They complement CASREACT, which is much larger and most suitable for identifying a specific reaction. Since REACCS is an in-house system, it allows for unlimited searching and browsing without the accumulation of online charges.

ORAC is a software package that allows searching among chemical reactions by drawing reactants, reagents, solvents, and products using molecular structures and substructures and by using hierarchical keywords and a variety of other data types. It can be used to search simultaneously proprietary and commercially available databases.

Physical and Chemical Properties Databases

One of the more frustrating and time-consuming tasks facing chemists is the search for physical or chemical property information about specific materials. Several problems are associated with such searches. First, the data are scattered throughout a vast literature. Second, any collection of measurements of a specific property by different researchers at different times will have inconsistencies that must be taken into account in the search for the most accurate numbers.

The advent of computerized literature searching has given rise to the hope that a satisfactory solution to these vexing problems could be found. In fact, the development of numeric and property databases has lagged far behind the rapid growth of bibliographic files, and only recently have numeric and property databases begun to appear.

Over the past two years, STN has significantly expanded its coverage in this area and now offers ten databases covering a broad range of materials and properties; more are promised. As an aid to users coping with this proliferation of databases, STN has also developed a thesaurus and data directory, the NUMERIGUIDE File, that serves as a convenient and inexpensive guide for the online searcher to use to determine which databases contain the property information sought. Preeminent among these files in terms of size and scope is Beilstein. A sampling of some of the smaller, more specialized files offered by STN includes the following: DIPPR File, which contains physical property data for 1,117 commercially important substances; HODOC, which provides the most frequently used physical and chemical data for over 25,000 organic compounds; ICSD, which is a comprehensive file containing complete structural data for 29,000 inorganic compounds; and JANAF and NBSTHERMO, which contain compilations of thermodynamic properties.

STN's new service, the Materials Property Data Network, will incorporate seven new databases containing property data for classes of materials such as steels, aluminum and its alloys, ferrous and non-ferrous metals, high-temperature metals, and polymers.

Other commercial vendors also provide access to numeric and property databases, including Beilstein.

Finally, specialized collections of property data are commercially available on CD-ROMs, diskette, and tape, offering access to data such as mass spectral compilations, crystal structure data, and thermochemical values.

Chemical Information Management Systems

MACCS-II, an integrated chemical information management system produced by MDL for in-house databases, provides for searching by molecular structures and substructures. Thus, it gives researchers the same capabilities with respect to their own databases as those offered by the CAS Online databases.

OSAC is an integrated chemical information management system for in-house use. It provides for searching by molecular structures and substructures and by hierarchical keywords and a variety of other data types. Multiple database access is supported.

Front Ends to Commercial Databases

Software packages allowing users to build text or structure queries offline with the aid of extra search, display, and help features include Chem Talk Plus by MDL for IBM-compatible personal computers, STN Express for IBM-compatible and Macintosh personal computers, and Soft Shell products for the Macintosh. Alchemy II displays chemical structures in three dimensions and is designed to be used in conjunction with STN Express.

Future Needs and Directions

Serial and Patent Literature

Quick access to primary literature, both current and retrospective, will continue to be a fundamental need. More than ever, chemists need to go beyond the department library to obtain the information they require. There are several corollaries to this:

- Bibliographic control over primary chemistry literature, already excellent, will remain critically important.
- In addition to bibliographic control, chemists need easy and quick access to the documents themselves, especially to those not held by their own library, through efficient and inexpensive document delivery systems.
- With the increase in the quantity of chemical information, chemists frequently complain of information overload and a need for methods of interpreting or filtering out information that is not relevant. As artificial intelligence systems develop this capability, it should be applied to chemical information systems.
- It is important to ensure access to older chemistry journals that are deteriorating. Generally, preservation of the intellectual content rather than the physical document will suffice.

Integrated Information Environment

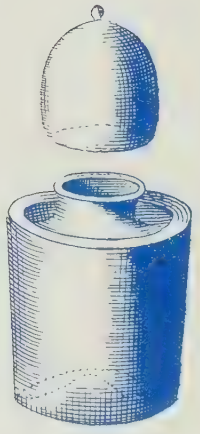
As increased amounts of information are available in electronic form, an integrated information environment for chemistry will become possible. In its most advanced form, this will involve:

- Increased use of national and international computer networks for communication among researchers, access to specialized sources of chemical information, and submission of articles and grant proposals.
- Local area networks containing the full texts of key journals (with graphics), important handbooks, directories and encyclopedias, and gateways to commercial databases.
- Interfaces with laboratory information systems and their files of numerical data.
- Large software collections on a local mainframe computer.

In short, chemists envision the ability to communicate with colleagues via electronic mail, retrieve the text of an article through a gateway, download information into a word-processing file, import structures drawn in a graphics program, add scanned spectra, and send completed articles to a journal by electronic mail—all from their own terminal. The Chemist's Crystal Ball, a project begun in 1990 at the Yale University Libraries, will develop a model of such an integrated electronic information environment. Plans are to provide access to a wide range of electronic information, including bibliographic databases, handbooks of physical data, bulletin boards, journals, preprints, and crystallographic databases.

Competing Needs

It is clear that information requirements are growing and becoming more expensive. Journal subscription prices are accelerating at a rate that far exceeds that of inflation, new journals continue to proliferate, and researchers are becoming increasingly dependent on costly online services. At the same time, chemistry facilities in universities are in dire need of renewal, and laboratories must keep pace with new technologies. Budgets, of course, are finite and in some cases, shrinking. In the future, hard choices will be necessary to reconcile the conflicting demands on resources.



Biology

General Characteristics

Nature of Research

Biology, the study of living things, is an extremely diverse and varied discipline, described by biologists themselves as "messy." This is a feature that appeals to many biologists, who delight in devoting their energies to discovering biology's organizing principles. The discipline's heterogeneity also makes it difficult to describe biology as a coherent whole.

Biology over the past few decades has entered into a golden age of expansion and discovery. Many would attribute this trend to the development of a model for DNA in 1953. In the past 10 to 15 years, this has led to the development of recombinant DNA techniques, which have been applied in virtually every subfield of biology. It has also led to the breakdown of traditional subfields within the discipline and the advent of a more integrated approach that emphasizes the questions being asked rather than the type of organisms being studied.

During this period, biology has evolved from a discipline that was largely descriptive to one that also emphasizes experimentation and analysis. Although no hard and fast rules separate the different types of biology research, it can be said that the organismal, natural history side of biology usually employs the descriptive approach, while the molecular, functional, and chemical side of biology generally uses the experimental approach. Although overlap between the two approaches exists in some areas of study, e.g., ecology and evolutionary biology, the research methods and interests of each approach are in general very distinct. However, some biologists predict that the application of approaches from molecular biology to all areas of biology will result in greater convergence between the experimental and descriptive methods in the future.

Some fields of biology, like population genetics and certain subdisciplines of ecology, make extensive use of mathematics, and almost all biology research makes use of statistical analysis.

Biology is increasingly equipment intensive, with advances in computers and analytical equipment such as DNA synthesizers and spectrophotometers fueling more complex lines of inquiry and further discoveries. This move toward increased technical complexity, however, makes biology less easily understood by laypeople.

Biology has a strong influence on other fields. It is an integral part of research in agriculture, medicine, and the environment. Its molecular side intersects with chemistry and physics; its organismal side connects with anthropology, paleontology, and psychology.

Nature of Information

The body of biology research is extremely large and varied. Unlike chemistry information, much of which can be organized around molecular structures, biology information has multiple centers. Probably because biology is so diverse, it also lacks chemistry's long tradition of well-controlled literature.

Like research in other science disciplines, research in biology is impossible without reference to previous research.

The Future for Biological Information

As biology expands and becomes more complex, the information needs of biologists will follow suit.

Biology is the basis of several growing industries which invest heavily in basic research. To remain competitive, they must have access to the latest technology and information.

Use of Biological Information

As in chemistry, biologists seek to answer specific questions, to gain in-depth background knowledge in a narrow area, and to keep current with trends in a broader area. But because biology is so diverse and heterogeneous, the questions biologists ask and the ways they approach their research are extremely varied.

A basic question all ask at some point is, what work on a subject has been done before and by whom? As in other disciplines, questions vary from general to specific and might include: What homologous processes, structures, enzymes, and genes exist in other organisms (than the one being studied) and what has been the result of their study? Has the three-dimensional structure of a particular protein been described? What sequence information exists for a particular gene and what papers have been published on it? What work has been done on a particular genus or species in a particular geographic location? What is the geographic distribution of a particular species? What does a particular substance do in the human body? Has anyone cloned this gene previously? Has anyone patented this process? What is the best procedure for growing a particular strain of yeast, or some other organism?

Online bibliographic databases allow biologists to search the last 20 years of literature for answers to these questions, using keywords and names of particular organisms, enzymes, or diseases as access points.

Primary Literature

Serial Literature

Because of its reliance on experimentation and observation, biology depends heavily on reports of previous research in the periodical literature. Most biologists regularly read six to ten journals from cover to cover, often including *Nature* and *Science*, which report on important recent developments across the science disciplines. Organismal biologists may browse as many as 100 journals. Keeping current with the literature is increasingly difficult, as the number of journals to contend with seems to grow exponentially.

Reflecting changes in the discipline, the nature of biology literature has changed over the past 30 years. Formerly, articles consisted of text accompanied by line drawings or photographs, which were generally accessible to the educated reader. Now, a smaller quantity of text is accompanied by a large quantity of formulae, charts, tables, color photographs, and complex illustrations that are difficult for the layperson to comprehend. Along with these changes have come increases in the price of journals, in part because production has become more complex. In some cases, prices are approaching those of chemistry journals, which are among the highest in the sciences.

Access to journal literature is almost entirely through print publications. Some observers have noticed a ground swell of interest in electronic journals distributed either online or offline, on media such as diskette or CD-ROM, and at least one group, the American Society for Microbiology, has launched an experiment with publication in both print and CD format. However, observers predict that before electronic publishing can gain a hold, several hurdles must be surmounted: the difficulty of transmitting color photographs and other graphics, the absence of standards for encoding and distributing texts and accompanying images, and library and user reluctance to forgo paper copy. The electronic versions of journals that are published in both print and electronic formats may have difficulty succeeding because publishers incur substantial additional costs for little economic benefit. Finally, publication of electronic journals is not necessarily faster than that of conventional journals, since the most time-consuming phase of the publication process—the work performed by editors and reviewers—is affected little if at all by automation.

Patent Literature

Worldwide patent literature has long been vitally important to research in the corporate arena but in the past was used infrequently in academic research. However, biology librarians say that this pattern is changing and that they are beginning to field more and more patent queries, particularly in such fields as agriculture, immunology, biotechnology, diagnostics, and antibiotics.

Major Indexing and Abstracting Services

Because biology is so broad and varied, no single indexing and abstracting service comprehensively covers the entire discipline. The most frequently used are *Science Citation Index*, *Biological Abstracts*, *Index Medicus*, and *Zoological Record*. For chemistry-related research, biologists also consult *Chemical Abstracts* (see **Chemistry**).

Print Indexes

Science Citation Index (*SCI*) indexes some 1,200 life science journals worldwide, or about 246,000 articles every year, in addition to journals in

other science fields, technology, and the behavioral sciences. When cited, patents are also included. The number of journals covered, although smaller than the number covered by *Biological Abstracts*, represents what the producers describe as core life science journals, defined as those most often cited. *SCI* is unique among indexes in enabling the researcher to trace work that was influenced by a seminal article—in other words, by following a citation trail the researcher can trace research forward in time and identify work grounded in an earlier article. Used properly, the citation method can pinpoint citations that subject or keyword searches might never turn up. *SCI* is published bimonthly and is also available on CD-ROM.

Biological Abstracts monitors approximately 9,000 sources worldwide, providing coverage of virtually every aspect of the life sciences. Published since 1927, it is arranged by subject with additional author, keyword, biosystematic, and generic indexes. The two latter are particularly useful to biologists, the first used to find entries by taxonomic category (e.g., insects, amphibia, mollusks), the second used to find entries according to genus or genus-species names (e.g., a specific virus or amphibian). *Biological Abstracts/RRM*, a complement to *Biological Abstracts*, provides semimonthly coverage of reports, reviews, and meetings in the life sciences.

Index Medicus provides references to articles from approximately 3,000 of the world's biomedical journals, covering both basic as well as clinical research. Abstracts are available for about 60% of citations published after 1975. Issued monthly by the National Library of Medicine, *Index Medicus* has been in print under this and previous titles since 1879.

Established in 1864 by a group of biologists associated with the Zoological Society of London and the British Museum, *Zoological Record* is now published jointly by BIOSIS and the Zoological Society. It aims to provide exhaustive coverage of the literature of systematic zoology and covers about 5,500 serials worldwide. It is published annually in 27 separate sections; 25 deal with different animal groups (e.g., mammalia, amphibia), and the other two cover general zoological literature and new genera. In addition to subject and author indexes, it offers geographical, paleontological, and systematic indexes.

The literature of agricultural and allied sciences is indexed by the National Agricultural Library's *Bibliography of Agriculture*, which contains citations to journal literature, government documents, monographs, and conference proceedings. Fields covered include agricultural economics and production, plant science, entomology, food and human nutrition, forestry, natural resources, soils, and fertilizers.

Biology also has a host of indexes devoted to specialized areas of research. A sampling of these includes *Ecological Abstracts*, offering comprehensive coverage of current ecology literature, and *Biochemistry Abstracts*. Further, the Commonwealth Agricultural Bureaux (CAB) in Great Britain publish 31 journals with abstracts covering all areas of the

worldwide agricultural literature, including materials that are relevant to biological research. *Helminthological Abstracts* covers all aspects of human, animal, and plant parasitic worms.

Electronic Indexes

Biological Abstracts and *Biological Abstracts/RRM* became available online as BIOSIS Previews in 1969. The database contains more than 6 million citations with abstracts to the worldwide literature since 1969 on research in the life sciences, including microbiology, plant and animal science, experimental medicine, agriculture, pharmacology, ecology, biochemistry, bioengineering, and biophysics. In addition to information from periodicals, it also includes information on monographs, conference proceedings, research reports, and symposia. An attractive feature of the online database is the addition of biosystematic codes enabling the user to search by type of organism. In 1989, BIOSIS introduced BIOSIS Previews/RN, which includes searchable CAS Registry Numbers (see **Chemistry**) for approximately 1.4 million records.

MEDLINE, which corresponds to *Index Medicus* and in part to the *Index to Dental Literature* and the *International Nursing Index*, was made available online in 1966 and now contains more than 5 million records. It provides access to the worldwide biomedical literature published since that date, including research, clinical practice, administration, policy issues, and health care services. References are drawn from approximately 3,400 journals; abstracts are available for about 60% of the citations.

SCISEARCH, the online version of *Science Citation Index*, covers worldwide science literature published since 1974. Although its coverage is not as comprehensive in any one area as that of MEDLINE and BIOSIS Previews, researchers in biology frequently find it useful. Articles indexed in SCISEARCH can be obtained through ISI's document delivery service, The Genuine Article.

The *Bibliography of Agriculture* is available online as the database AGRICOLA, which provides access to agricultural literature since 1970 and contains some 3 million records.

Zoological Record Online corresponds to the print index and contains more than 700,000 citations to the worldwide literature on zoology published since 1978.

The online version of Commonwealth Agricultural Bureaux publications, CAB Abstracts, contains more than 2 million records covering the literature of the agricultural sciences and related areas of applied biology published since 1972.

All of the above are valuable tools; one drawback noted by researchers, which reflects their dependence on older literature as well as current, is the lack of retrospective coverage in the online indexes. Also, biologists say they find them difficult to search without the assistance of an expert.

For information about patents in biology, an important source is the BIOSIS-produced database BioBusiness. In addition to covering the worldwide periodical literature on business applications of biological and biomedical research, it also covers patents in such areas as immunological testing, food processes, and fishing.

GENESEQ, a database introduced in 1990 by Derwent Publications and IntelliGenetics, Inc, provides information on protein and nucleic acid sequences from published patent applications. Sequence information is gathered from patents filed in the U.S., Europe, Japan, and other major patent offices. In addition to the sequence itself, a GENESEQ entry contains descriptive material about the sequence. GENESEQ links with Derwent's other patent databases, including World Patents Index and Biotechnology Abstracts (see **Chemistry**).

Current Awareness

Biology has not escaped the information avalanche that other science fields are experiencing. An expanding field, it offers many opportunities for innovation and discovery and has attracted increasing numbers of scientists to its ranks. Because protocol in science dictates that the first person to announce a discovery is the one to receive credit for it, biologists are driven to report their research in the journal literature, the discipline's official record. The result is enormous and growing quantities of information. Although biologists may complain about information overload, the phenomenon also challenges them to keep current.

There are several means of doing this, each with advantages and drawbacks. Of greatest importance is the journal literature because it has generally undergone quality control through the refereeing process. But refereeing and the other parts of the publication process take time. The weeks and months they require can be frustrating to biologists because many fields are moving so quickly. The average time between submission of a paper and publication is between six and eight months; in some cases, the gap approaches one year. Conference proceedings are useful, although unrefereed and thus less reliable than the journal literature and not always easy to obtain. Informal networks are essential for testing ideas and fine tuning research methods, but access is less than universal.

Online Tools

The time lag between a paper's publication and the appearance of a citation and abstract in MEDLINE or BIOSIS Previews is at least six to eight weeks. SDI LINE and BioExpress, two other online services, offer bibliographic citations (without abstracts) to journals listed in those indexes within a week or two of their publication, thus helping to bridge the gap between publication and indexing.

Current Contents Search is a popular current awareness tool, often used in conjunction with a program that allows a researcher to construct a profile of his or her interests, which is run periodically against the database to extract the relevant records on floppy disk. Another option is Reference Update, an online service which offers, in addition to basic table-of-contents information, keyword access to recently published articles. The above services also offer document delivery.

Conferences and Informal Networks

Because collaboration is absolutely fundamental to biology research, biologists rely on sources other than bibliographic databases for keeping current. Symposia, like the Cold Spring Harbor and Gordon conferences, bring together researchers from around the world to focus on one topic. They are a stimulus to creativity and a source of up-to-the-minute information about current research projects and experimental techniques. But conference proceedings have generally not undergone review and may contain incomplete or unverified information, so researchers use them with caution.

Some biology subfields enjoy the benefit of informal, unrefereed newsletters which disseminate information of general interest within the community. An example is the *Drosophila Information Service*, established in 1934 to promote communication and prevent duplication of effort among the growing numbers of *Drosophila* researchers. It contains information on widely used genetic techniques, *Drosophila* stocks, technical notes, and general news and announcements. Beginning in 1991, it will be distributed online.

When the number of researchers working in the same field is small, the telephone is often the most useful tool for exchanging research results before they appear in print. Increasingly, biologists are also using electronic networks to communicate with colleagues. In molecular biology, it may take several months, if not longer, for a protein sequence to be entered into a central database such as GenBank (see below) and up to two years to appear in the literature. Researchers will take a short cut by using electronic mail or a telephone call to contact colleagues who are working on a particular sequence to ask them to send the results when they are complete. Use of electronic networks for communication and exchange of data is facilitated by government support for the development of databases and computer networks, particularly in the area of genome research.

Collections of Data

Increasingly, biologists rely not just on data they collect themselves but also on data others have amassed. They can do this because advances in computer technology make it possible to organize research results so they are communally accessible to a degree not attainable before. In fact, the trend toward making collections of data available through computer networks is opening up new lines of inquiry in biology. Several initiatives described below illustrate this trend.

Museum Collections

Researchers in the subfields of biology related to natural history and the environment depend heavily on collections of specimens. Natural history museums, academies of science, herbaria, and similar institutions maintain collections of plant and animal specimens which can number in the millions and usually focus on specific geographical regions or on categories of flora and fauna.

Biologists have long used taxonomic categories (which place an organism within a particular phylum, class, and genus) as a basic structure for organizing information about plant and animal specimens and natural his-

tory collections themselves. But information about where a specimen was found (e.g., Mount Day, Mount Hamilton Range, altitude 3900', Santa Clara County), when it was found, and the name of the person who collected it, which is vital for studying the distribution of flora and fauna, might only be obtained by going through a shelf of specimens and examining the labels one by one.

Outside the repository, information about these collections is at best very general. For example, a researcher might learn that a herbarium specializes in the flora of California, Nevada, and the Andes, but information about specific holdings is not available from union catalogs analogous to library bibliographic databases (although some collections maintain local databases for in-house use). The demand for this information goes beyond the academic biology community to state and federal conservation agencies, local planning agencies, and private environmental consultants.

Within the natural history community, awareness of the need for central databases where this information can be shared and for standards that allow databases to be used by a variety of communities and disciplines is growing. The Association of Systematics Collections (ASC), representing the natural history community, has established the Task Force on the Computerization and Networking of Natural History Collections to address requirements for databases, coordinate efforts to collect information about specimens, and reach agreement on standards for describing them.

A project being developed by the Association of California Herbaria, representing 7 million specimens in 145 California herbaria and working collections, will lead the way. The organization plans to develop a database of 500,000 botanical specimens from California—in effect, an inventory of the state's flora, including its rare and endangered species. It will accommodate queries that are virtually impossible to answer now: What taxa have been documented to occur in a specific geographical area? What is the distribution of a specific taxon? What is the range in elevation for a specific taxon? What specimens were collected by a particular individual?

Biologists involved with ASC initiatives believe that many benefits will accrue from the move toward automation and networking. Access to collections will be improved and accession policies will be more easily coordinated among them. Geographical areas that are sparsely represented will be more easily identified. Detailed studies of correlations between distribution and habitat factors will be possible. The distribution of rare and endangered species will be more accurately determined. Finally, it will be possible to connect to databases with information from other regions to examine the distribution of taxa that are found in a number of geographical areas.

Other Data Collections

Over the past 15 years, automation has spread rapidly in the microbiology community. Its adoption has underlined the need to integrate separate, local databases into resources that will be widely useful and available.

Although increasing numbers of scientists are developing in-house data sets, many of them cannot be easily shared because their "homegrown" characteristics prevent others from adapting them. Researchers are not always assiduous about documenting files developed for their own use. Further, they may organize the data in a way that suits their own research needs but does not lend itself to other uses. Since creators of data sets receive little or no benefit from reformatting and documenting the file for use by others, they may not make this effort. As a result, further use is often limited. However, interest in facilitating the sharing of data that are collected at great expense and effort is growing.

Two major efforts are under way. The National Library of Medicine's National Center for Biotechnology Information is working to solve some of the problems described above as they relate to biotechnology by creating automated systems for information on molecular biology, biochemistry, and genetics and coordinating efforts to gather biotechnology information worldwide. Also, the Human Genome Program at the National Institutes of Health is leading an international initiative to characterize all human genetic material and to make this information available to the scientific community.

The activities of GenBank On-line Service (GOS; formerly Bionet) illustrate the complexity of the task, which involves communication with researchers who generate information, computers to store it, software to analyze it, staff to maintain the databases, and international computer networks to gain access to the data. GOS provides access to biological sequence databases, software tools to analyze the data, and electronic communication among scientists. The service allows researchers to identify genetic patterns, detect similarities in structure, and plan experiments using the most up-to-date information available.

One database available through GOS is GenBank, which contains descriptions of the chemical and physical makeup of DNA and RNA sequences reported since 1967, as well as bibliographic citations to articles describing nucleotide sequences. There are a number of access points to the database, including nucleotide sequence, organism name, gene name, author, and accession number. To promote data sharing, which is now required by most funding agencies, GOS also offers a facility for submitting newly discovered sequences to IntelliGenetics, Inc., a coproducer of the GENESEQ database.

An important element in the development of electronic databases and related services is the bridge they can provide to the biological literature. The National Library of Medicine is developing a common search vocabulary which retrieves both genetic sequence records and the biomedical literature that references the same concepts. Already, linkages have been created between MEDLINE and GenBank using the NLM's Medical Subject Headings (MeSH) indexing vocabulary; links to other databases are also under development. These links between the data and the literature are likely to produce increased demand for the literature, which will have a major impact on libraries.

Government agencies like the Department of Agriculture, National Climatic Data Center, and Forest Service also produce data of use to biologists. Since each agency has a different system for providing information about and ordering data sets, locating them can be a challenge. In the case of both government and private data sets, the publication and distribution system that has evolved for print materials does not suit computer files well. Few are published through the normal channels. Also, the system of reviewing that alerts librarians and researchers to developments in print does not exist for scientific data sets.

Once a computer file is obtained, the ease with which it can be used also varies. Many arrive without the enhancements, such as software for analysis of the data, that the publication process customarily provides. In some cases, for-profit information vendors offer these enhancements but at a significantly higher price.

Standards

The development of databases has given impetus to the issue of standards for data description. Because researchers often gather information from many sources, researchers and funding agencies are increasingly aware that standards suitable for interdisciplinary, international research are imperative.

Future Needs and Directions

Biology Literature

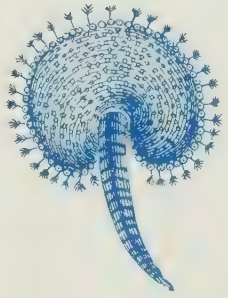
Biologists will need a broad array of materials from both biology and its many related areas of research. Because of the need for evaluated data, methods, and procedures, journals will continue to be the principal medium of communication in the discipline. Several factors will affect access to and use of the literature:

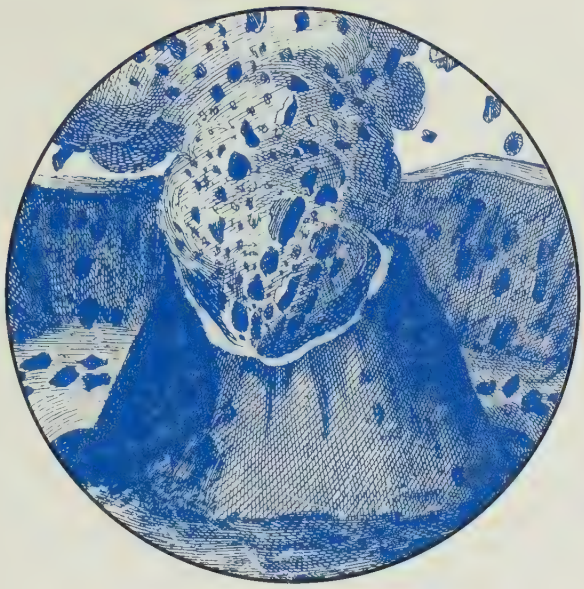
- Research reports published before the mid–1960s will continue to be important; more retrospective coverage in electronic indexes is needed to improve access to them.
- The rapidly rising cost of journals will probably force libraries to collect fewer titles than in the past and to rely on outside sources for access to the entire array of information needed.
- The need and demand for links between biological data and literature, whether through registry numbers or a common search vocabulary, is likely to grow.
- Experiments with electronic journals may become more common. But until it is feasible to transmit color photographs and other graphics, biologists will use them principally for full-text searching and will still require access to a print copy.
- Keeping current will become more and more difficult as the biological literature continues to grow. Both librarians and researchers will continue to be challenged by the breadth and diversity—the "messiness"—of the field.

Data Exchange

The organization of research results, whether information on endangered species or genetic sequences, so that they are communally accessible will become increasingly vital, and will depend on:

- Development of standards for data description; in many areas, the lack of standards stands in the way of data exchange.
- Specialized software for analysis of data.
- Coordinated efforts to collect biology data worldwide.
- Better access to existing life science data sets.
- Widely accessible networks with speeds sufficient to exchange large quantities of information.
- More services that bring related resources together and thus make access easier.





Geosciences

General Characteristics

Nature of Research

The geosciences study the earth's systems—solid earth, atmospheres, and oceans—and their interaction. The six principal geoscience fields are geology, paleontology, geochemistry, geophysics, hydrogeology, and climatology. The last four are newer fields than geology and paleontology and easier to define.

Geoscience research is often done in pursuit of natural resources or, more recently, to protect natural resources.

Increasingly, as scientists seek to understand the interdependence of the earth's natural systems, geoscience research no longer follows the lines of classical disciplinary boundaries: rocks and other terrestrial objects are now often viewed as chemical systems subject to physical laws and modified by the living biosphere. Observers predict that in the future, areas of study with labels like "biogeochemical dynamics" will become common.

Nature of Information

With the exception of theoretical information, most geoscience information is site and/or time specific: it refers to a particular spot in, on, or above the earth and relates to a particular epoch. This characteristic affects the way information is organized and located.

The information resources required to support geoscience research are vast and varied. For example, maps still provide the vital entry point to the researcher in the field; numeric data are important to the geophysicist; and the laboratory is imperative to the geochemist.

Geoscience has a strongly historical orientation. Old observational data are still useful even if in many cases the methods of observation or analysis have changed.

Geoscientists frequently use information from geographically remote locations. When geoscience information collected in developing nations is published locally, distribution may be erratic. This makes it difficult for libraries, whether local or abroad, to get the information.

Nearly every federal government and many state- and provincial-level governments worldwide support geological research. The publications of government-supported geological survey offices are indispensable to geoscience research. However, a significant amount of geoscience data is proprietary, gathered by profit-making companies or consultants. If they make data available to outside researchers, they charge a fee for doing so.

The Future for Geoscience Information

Huge quantities of data collected by remote sensing instruments on platforms such as satellites and ships are transforming research in some geoscience fields. The large amounts of data now available, along with tools for their analysis, allow geoscientists to develop new explanations for natural phenomena (e.g., the theory of plate tectonics).

The geosciences are a focal point for research tracking changes in global conditions: warming and cooling events, migration of continents, rising and falling sea levels, variations in floral and faunal conditions, and the appearance and disappearance of deserts, marshes, and mountains.

The quantity of geoscience information, already huge, is growing at an accelerating pace. Access to this information will be critically important and will depend on increasingly sophisticated systems to provide both access and the capacity to manipulate the information found.

Use of Geoscience Information

Geoscience research is extremely varied, and as a consequence geoscientists employ different methods of attacking a research problem. One might start with a rock sample, another with an environment, and yet another with a mineral deposit. Some might start by browsing in the current literature.

No single paradigm for geoscience research can do justice to the discipline's complexity, but it is useful to look at how a geoscientist looking at a specific geographic area might proceed. First, he/she looks for maps of the area—topographic, general geologic, or tectonic; or, more specifically, hydrologic, geochemical, and so on. Next, the researcher would look for field guides covering the area. Following that, he/she might perform searches in indexing and abstracting tools and add the results to citations found in the field guides. Finally, the researcher would move on to more specialized sources. The goal is to collect as much information as possible about the area.

Questions geoscientists might ask range from the broad to the specific. What information has been published on the geology of a particular area? What maps are available for a particular area? What information has been published on the geology of a particular stratigraphic unit? What are the locations of the major sources of a particular mineral? What is the commercial output of a particular mineral for a certain year in a particular

country? What are the physical, optical, and/or chemical properties of a given mineral? What are the latest results on a particular geochemical reference sample? What descriptive and high-quality photographs of a particular fossil taxon have been published?

Primary Literature

Assembling the broad selection of materials required to answer the above questions is a demanding task. The sheer number of titles alone required for a research collection poses a challenge. Further, by comparison with other disciplines, geoscience materials are very diverse with respect to date and place of publication, subject, the variety of formats, and types of information sources.

Because the information needed for geoscience research is so broad and varied, informal approaches to the realm of research information are particularly prevalent. As in most other fields, personal subscriptions, personal networks, conference proceedings, literature reviews, exchange of preprints, and browsing in the library are essential for keeping up with recent research.

Serial Literature

As in the other science disciplines, journals contain the most recent evaluated information and are the most prestigious publication outlets for academic researchers. Although geoscientists' reliance on journals is not quite as heavy as that of physicists and chemists, serials are nonetheless the single most frequently cited source of information. With the selective cancellation of library subscriptions, more noticeable in recent years due to sharp journal price increases, geoscientists and librarians worry that academic libraries will be unable to maintain the breadth of coverage necessary for specialized research.

The breadth needed varies widely. One seismologist estimates that 90% of what he needs is published in four journals. But for others, browsing often ranges far beyond what is typical in other fields. One researcher claims that combing more than 100 journals systematically is not atypical and that the literature is so vast it is impossible to cover it all. One geologist regularly browses 150 journals and assembles citations in a file covering 20 years of geological literature. A mineralogist would likely find half of the literature he/she needs in the physics, chemistry, and materials science fields.

Monographs

Monographs usually contain conference proceedings or compilations of papers. They may also be texts containing information digested from the journal literature. Even though they are less valuable than serials for current awareness and lack the detailed regional information found in theses, dissertations, and guidebooks, monographs are an important source of information, particularly for historical background.

Government Documents

Nearly every federal government and many state- and provincial-level governments worldwide support geological research through geological surveys, bureaus of mines, departments of natural resources, and so on. The publications of these bodies—maps, guidebooks, newsletters, journals, reports—are indispensable to geoscience research. Locating and

collecting these publications is difficult: production is uneven, communication with survey offices in remote countries is often troublesome, and the customary distributors of serials and monographs do not carry government publications. Earth science libraries devote considerable resources to collecting them, and some have amassed large collections focusing on certain regions.

U.S., Canadian, and Australian government surveys, both national and provincial or state, also issue open file reports—unrefereed reports that have not been formally published. Some U.S. Geological Survey reports, issued on microfiche, are indexed by the major indexing services, but state and provincial survey reports are not indexed at all, are often unavailable on microfiche, and are extremely difficult to obtain. Some earth science libraries collect them for particular regions; to acquire reports outside their own collection, librarians must rely on their knowledge of other libraries' collecting policies or enter the government agency labyrinth.

Maps

Maps, critical to geoscience research, are published in many ways: separately as monographs, in series, and bound or folded into a pocket in the back of a text. Most geologic maps are printed in color and are much larger than a standard sheet of paper. Geoscientists generally need the original printed copy (and will continue to do so until the availability of large-copy color copiers) to carry into the field with them.

Maps pose a variety of other problems. Bibliographic control is poor, varying widely from collection to collection. They are notoriously difficult to obtain through interlibrary loan; sending them through the mail is difficult because of their size; photocopies are useless because maps are so large and require color; and microfilms are expensive.

Preservation is also a concern: many are printed on acid paper, which deteriorates over time. In addition, they often suffer wear and tear when taken into the field. Because maps are generally not bound, they tend to be pilfered from library collections. Finally, computer-generated maps may be lost when they are not saved in hard copy form or when they are saved on magnetic media that may become obsolete.

Professional Society Publications

In the geosciences, national, state, and regional professional organizations have a long tradition of publishing research-level serials, monographs, maps, and field guidebooks. Examples of societies include the American Association of Petroleum Geologists, the New Mexico Geological Society, and the Midwest Friends of the Pleistocene. The physical quality of their publications varies widely. Sales and distribution efforts are erratic, and publications are rarely available through commercial vendors; moreover, most societies do not accept standing orders. Some societies even lack a permanent address. As a consequence, professional society publications are among the most difficult of geoscience materials to obtain.

Guidebooks, Theses, and Dissertations

Geologists especially value guidebooks, theses, and dissertations for their detailed descriptions and analyses of outcrops and features—in effect, the original data sets of geoscience research.

Guidebooks, often issued in conjunction with geologic field trips, contain detail generally unavailable elsewhere: road logs, outcrop descriptions, proprietary information from industry, and supporting text in which the authors synthesize current research. Guidebooks are very uneven in physical format, editorial conventions, and availability and distribution and are among the most difficult of geoscience materials to identify and obtain. Bibliographic control of guidebooks has long been inadequate but has improved significantly with the publication of the fifth edition of the Geoscience Information Society's *Union List of Geologic Field Trip Guidebooks of North America*. The indexes, which were difficult to use in earlier editions, have also improved. Conference proceedings are another important source of guidebooks.

Many geoscience theses and dissertations involve fieldwork along with a survey of the area's geologic history and relevant published research. They are a very important resource, especially since the detail they contain is generally not published elsewhere.

Many doctoral dissertations are abstracted by Dissertations Abstracts International and indexed in the main indexes. But index coverage of master's theses is much less comprehensive. Moreover, few dissertations and theses are cataloged on national bibliographic utilities. Even if a copy can be located, it may not be obtainable.

The paper on which dissertations and theses are written is in many cases deteriorating, and the maps and other materials (e.g., stratigraphic charts and plot diagrams) that accompany the work, usually in a pocket, are often in even worse condition. Many university libraries, charged with preservation responsibility for what is often a unique copy, refuse interlibrary loan requests. Microfilm copies are of limited use, since thesis maps are often hand-drawn in color.

Older Materials

Older serials, maps, reports, guidebooks, and theses contain valuable observations. Even if the earlier analyses are no longer valid, the facts these works contain retain their value. Egyptian flood control reports from the 1930s, for example, are used to provide data for dam construction; Soviet geologic surveys dating from the 1930s, because they describe natural seismic activity in the years preceding the development of the atomic bomb, can be used by scientists to derive a baseline for the purpose of nuclear test monitoring. But these materials are seldom cataloged in a bibliographic utility, thus limiting the likelihood they will be used by anyone other than local researchers. Also, many older materials have seriously deteriorated. In addition to better bibliographic control, preservation of older materials, preferably through some means of full-size facsimile reproduction rather than microform, needs attention.

Major Indexing and Abstracting Services

Because earth science research information needs range widely by both topic and language, no single index comparable to *Chemical Abstracts* for chemistry provides the breadth of information needed. For both current information and essential historical background, it is necessary to consult at least three indexes to get complete coverage of a subject. Two indexes

(discussed further below), the *Bibliography and Index of Geology* and *Referativnyi Zhurnal*, provide broad coverage of geoscience literature. But they must also be supplemented by one or more of the large number of specialized indexes.

Geoscience indexes go beyond the serial literature to include information on conference proceedings, dissertations, maps, and monographs. Two indexes provide retrospective coverage exceeding that of any other major science index: the *Bibliography and Index of Geology* covers literature published since 1785, and the *Bulletin Signalétique*, a French bibliography, begins coverage in 1750—testament to the importance of bibliographies to geoscience research and to the importance of older literature to the discipline.

Print Indexes

The *Bibliography and Index of Geology*, published monthly by the American Geological Institute, covers worldwide earth science literature. It succeeds the *Bibliography of North American Geology* (1785–1970), the *Bibliography and Index of Geology Exclusive of North America* (1933–1968), *Geophysical Abstracts* (1966–1971), and the *Bibliography of Theses in Geology* (1965–66). It includes books, serials, reports, maps, and North American theses, and is an indispensable research tool.

Referativnyi Zhurnal is a monthly abstracting journal published since 1956 by the Akademiia Nauk in the USSR. Of the journal's 12 sections, geoscientists mainly use two, "Geologiia" and "Geofizika." Entries are in Russian, although some bibliographic information is given in the language of publication; annual author and subject indexes are included. Despite the difficulty of using a Russian-language publication and the lack of cumulations, geoscientists praise *Referativnyi* as the most comprehensive and current of the geoscience indexes and as particularly useful for author searches.

Geoscientists also use discipline-specific indexes, a sampling of which illustrates the fragmented nature of geoscience literature: *Geological Abstracts*, *Mineralogical Abstracts*, *Bibliography of Fossil Vertebrates*, *Petroleum Abstracts*, *Selected Water Resources Abstracts*, *Oceanic Abstracts*, *Meteorology and Geoastrophysical Abstracts*, *Bibliography on Cold Regions Science and Technology*, *Oceanographic Literature Review*, *Zentralblatt für Mineralogie*, *Zentralblatt für Geologie und Palaeontology*, and the *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*.

Electronic Indexes

The online database GeoRef contains about 1.4 million citations and corresponds to the print *Bibliography and Index of Geology* and its predecessors. The publication *Bibliography and Index of Micropaleontology* (1972–present) has also been added to the online file. Unlike other online science indexes, GeoRef provides full retrospective coverage of the literature covered by the print indexes named above, going back as far as 1785 in the case of the *Bibliography and Index of North American Geology*. GeoRef offers searching by keyword and by latitude and longitude, a

significant enhancement not available in the print edition. However, geology librarians state that information is difficult to retrieve, especially for those without training. GeoRef is also available on CD-ROM.

Geobase, the online version of *Geological Abstracts*, contains more than 300,000 citations with abstracts to the worldwide literature on the earth sciences published since 1980. Sources include journals, books, conference proceedings, and reports. It is produced by Elsevier/Geo Abstracts.

Bulletin Signalétique is available online as Pascal:Geode, a bibliographic database containing more than 900,000 citations with abstracts to the worldwide earth sciences literature. Produced by the Centre National de la Recherche Scientifique, it provides coverage of books, periodicals, theses, reports, and conference proceedings. Literature concerning France extends back to 1750; coverage of world literature begins in 1968.

Geoscientists also use databases focused on the literature of other disciplines. These include CA, Compendex, and the online versions of *Zoological Record* and *Physics Briefs*.

Gaps in Indexes and Bibliographic Utilities

Even with this abundance of indexes, researchers estimate that a significant percentage of the relevant material, both current and retrospective, is not indexed at all. Some of the literature not indexed is, admittedly, highly specialized. A paleontologist whose specialty is flowering plants, a field of study covering over 3 billion years, finds indexes of little use. However, much of the literature omitted from indexes is required by a broad base of users.

Government documents—federal, provincial, or state—represent the most glaring gap in index and bibliographic utility coverage. At present, librarians rely on informal knowledge of other libraries' collecting policies, but more systematic access is needed. Open file reports are also inadequately represented. Although GeoRef provides references to federal open file reports, it does not index state reports, many of which are consequently lost (particularly since a significant number are never published). Finally, maps are also inadequately represented in indexes and bibliographic utilities. Many thesis maps are not cataloged at all; access to maps from other sources varies.

Other Electronic Resources

In the past, geoscientists' research was largely based on their own observational data or on those of close associates or collaborators. Even though most data had uses beyond the purpose for which they were collected, data sharing was relatively rare; when it was accomplished, it was done between individuals.

Today, with the emphasis in many areas of research shifting from observation to analysis, few geoscientists can do effective research without relying on data collected by others. Increasingly, geoscience research requires bringing together many different types of data from wide geographical regions. Geoscientists also need access to both recent and

historical data, sometimes covering long time spans. For example, a geologist studying African water resources may consult camel caravan charts as well as aerial photographs.

Much of the currently generated information is now produced in machine-readable form. But the bewildering array of data being collected presents problems of its own. Faced with a torrent of information, researchers are equipped with few means to determine if the information they need exists and, if it does, where it can be located. For instance, researchers now know that the hole in the Antarctic ozone layer discovered in 1986 was in fact recorded ten years earlier by a satellite, but the tapes had lain unused in a vast NASA archive.

The problem promises to become more acute. According to recent estimates, spacecraft will produce between 1990 and 1995 as much information as they did in the previous 20 years. After 1995, the volume will double every two years. As the quantity of information grows, questions and difficulties surrounding the management of it also mount.

Government Data Sources

Data sources include government agencies, universities and research institutions both in the U.S. and abroad, international organizations, and private industry. Government agencies are particularly active in collecting, archiving, and distributing both older and recent data.

The National Oceanographic and Atmospheric Administration (NOAA) is a U.S. government agency with a large information infrastructure called the National Environmental Satellite, Data, and Information Service (NESDIS). Under its umbrella, three data centers have been established: the National Oceanographic Data Center (NODC), the National Climatic Data Center (NCDC), and the National Geophysical Data Center (NGDC).

The array of information they provide is staggering. At NGDC, for example, data services are divided into 62 different topic areas ranging from paleoclimatology to tsunami. Information about its holdings can be found in announcements of recent data acquisitions, maps, and slide sets; and various directories with general information, including lists of contacts.

Recognizing the difficulties researchers face in sifting through listings of data sets from one or more agencies, NOAA established the National Environmental Data Referral Service (NEDRES), which provides general information about environmental data. Available through the information vendor BRS, NEDRES contains descriptions of more than 17,600 publicly available environmental data resources, including computer files, print publications, and organizations that provide environmental data. Sources include NOAA's own data centers, as well as other federal, state, and private organizations. Useful though this service is, NEDRES coverage is not sufficiently comprehensive or detailed for many researchers.

Beyond this effort, NOAA has more recently begun work on a central directory system that will include all of its data holdings. Rather than

attempting to catalog discrete data sets, the directory will provide pointers to the location of broad categories of data. NOAA plans to cooperate with the National Aeronautics and Space Administration (NASA) and the U.S. Geological Survey (USGS), which also amass large quantities of data, to exchange information and make the directory as comprehensive as possible. The agency estimates that completion is five years away.

Activities of NASA in the arena of geoinformation are also extensive. In 1966 it established the National Space Science Data Center (NSSDC), which collects, archives, and disseminates scientific data obtained from spacecraft and ground-based observations from sources both within and outside NASA. The NSSDC maintains a data archive and distributes certain databases online. The NASA Master Directory is an online search system providing brief overview information about NASA and important non-NASA space and earth science data, including data from NOAA. The directory, which is available through a dial-up connection, also offers network connections to catalogs or information systems where more detailed information can be found.

USGS is another major source of earth science data. As in the case of NOAA and NASA, the USGS houses a vast array of data resources. The data or, in some cases, information about them, are divided among a multitude of in-house information systems, each reliant on different hardware, database management systems, and user interfaces. For centralized access to information about its holdings, USGS has recently developed the Earth Science Data Directory, which covers major USGS data sets. But because of the size of the agency and breadth of its activities, many data sets it generates are not listed in the directory. The USGS also operates EROS, an archival data center for satellite imagery collected by a number of agencies including NASA and NOAA. EROS holds some 6 million frames of photography, all listed in an in-house database according to geographical location. Any interested individual can contact EROS and obtain a listing of the imagery in a specific geographical area that is available for purchase.

Means of Locating Government Data

The above agencies also provide a multitude of print catalogs with detailed information about their holdings. However, directories themselves pose challenges. One is data set titles, which are often too generic to be useful (e.g., "Numerical Coefficients") or are meaningless acronyms. Compilers of different directories who seek to make a title more descriptive may each assign a different title to the same data set. Further, directories often follow their own format for describing data. (NASA has developed the Directory Interchange Format, but some experts say that it does not adequately accommodate the earth sciences.) As a consequence, users often find it impossible to know whether two data sets with similar descriptions are actually identical or different. Indexing is often rudimentary or nonexistent. Thus, directories provide only minimal navigational aids through archives that have been described as "tape landfills."

Although some of the organizations described above have in-house databases with indexing systems for specific uses or materials, they operate

independently with little coordination among them. In some cases, coverage overlaps; in others, gaps remain. Efforts to cooperate have recently achieved some success, notably NASA's Master Directory. These efforts are useful and represent an improvement over the chaos that preceded them.

Monographs and articles based on extensive data collections are no more effective at providing access to the data on which they are based. A geoscientist who publishes a survey of the Mid-Atlantic Ridge will spend millions of dollars (often provided by the government) to generate the data, but in the final report the data will be mentioned in a reference only. There is no assurance that the data are available through a government data center. Further, some journals do not allow the publication of data that an article is based on. Government classification of data as secret is another barrier that can arise.

An added complication is that earth science researchers, as in other disciplines, usually collect and structure data specifically for their own research needs with little or no attention to the documentation and standardization needed for wider distribution. This practice prevents others from putting the data to their own use.

Private Data Sources

Researchers usually understand that proprietary information will not be available to them without a quid pro quo, often an agreement to participate in profit-centered research. This often means that academics will either have to wait to obtain useful research data or forgo access to it altogether. For example, information gathered by oil companies usually will not be available until it is at least 10 years old. In rare instances, oil companies will donate reference data such as well logs (which help geologists identify physical characteristics of the rock strata surrounding a well) for research use. Some states require that proprietary data be deposited in a public repository, but they often allow 10 years or more to elapse before the requirement must be fulfilled. Other states do not have this requirement, with the result that much valuable information gathered by the profit-making sector will never be available for academic research.

The increasing prevalence of the trend to put public information into the private domain can also limit access to research materials. The Landsat satellites, which pioneered the photography of the earth from space, are a case in point. Since privatization of data distribution, prices for aerial photographs have risen dramatically. Further, use of Landsat material is limited to the purchaser, which prevents libraries from making this material available to patrons even if the library can afford it.

Future Needs and Directions

Geoscience Literature

Access to an exceptionally wide range of materials, widely dispersed in terms of time and place of publication, sources, and formats, will continue to be fundamentally important. A number of activities, many requiring cooperation among research libraries, will improve access to geoscience literature. They include:

- Support for retrospective conversion of bibliographic information on pre-1970s serial and monographic literature.
- More cataloging of maps on national bibliographic utilities.
- Retrospective conversion of records for theses and dissertations at major academic institutions and distribution of information on national bibliographic utilities, along with cooperative agreements to make them available through interlibrary loan.
- Better access to guidebooks, possibly through a national center which would acquire and distribute copies of every guidebook available, or through cooperative collection development; additionally, cataloging on national bibliographic utilities.
- Cooperative preservation of maps and selected older geoscience publications printed on deteriorating paper.
- Cooperative collection development and resource sharing of serials to prevent diminished access through serials cancellations.
- More cataloging of open file reports on national bibliographic utilities.
- Exploitation of technological advances as they make possible the inexpensive reproduction of maps and plates in color, with excellent quality graphics and half-tone photographs.

Geographic Information Systems

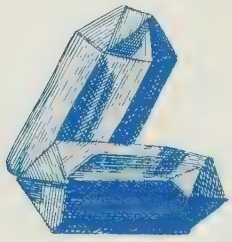
The directory and bibliographic methods of access to geoscience information are of limited use to a scientist looking for a broad range of information on a specific geographic location. Ideally, geoscience information should be organized by geographic location rather than by title, author, or name of a satellite.

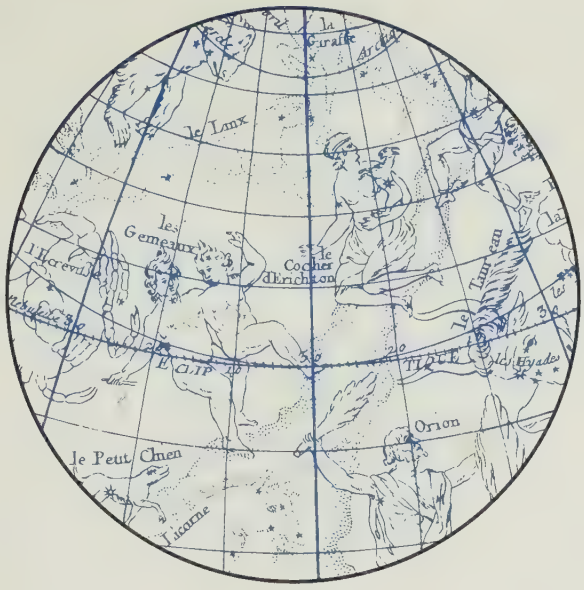
The geoscientist needs to be able to designate any given area or point on the earth's surface and, through this access point, unearth a wealth of information about the area. RLG has designed a system that would provide for centralized control and retrieval of the vast quantities of geoscience information. Called the Geo-Referenced Information Network (GRIN), this system would provide pointers to information as diverse as historical atlases and digital data recorded by satellites. The user only need draw a box around the area on the computer screen's map.

Beyond information about geoscience data, researchers also need systems that will actually manipulate the data. These geographic information systems would allow for the electronic storage and manipulation of large quantities of geographically referenced data. The design of such systems is a fast-moving new subfield. A 1989 grant from the National Science Foundation will support the creation of a National Center for Geographic Information and Analysis at the University of California at Santa Barbara; partners in the enterprise are the University of Maine and the State University of New York at Buffalo. It is expected that the new center will

provide a focal point for research in methods of displaying and storing geoscience data.

Another noteworthy project is Columbia's Geographic Access Program (GAP), which was established in 1986 to facilitate access to its huge data archive. It allows a researcher to select geophysical and geologic data for any area of the world, manipulate it, and perform calculations. As storage technologies and searching software improve and new data are added, GAP and similar projects will serve as increasingly powerful research tools for the geosciences.





Astronomy

General Characteristics

Nature of Research

Historically, astronomy has been concerned with the observation of the positions and movements of astronomical objects. More recently, its focus has shifted toward the study of the physical properties of those objects and what they reveal about how the universe began and may end.

The most important subdivisions within astronomy are determined by the type of object studied (e.g., sun, planets, galaxies). Within these subfields, research is usually done using a certain wavelength (e.g., most solar research uses x-rays).

With exciting new projects such as the space station either under development or under way, astronomy has entered what promises to be an age rich in discoveries never anticipated.

Ever since the Venetian Council granted Galileo a lifetime appointment to the University of Padua in recognition of his invention of the telescope, governments have provided significant support for astronomy research.

More than many other sciences astronomy is very data-dependent: to understand and identify new phenomena, theorists must have access to vast amounts of observational data.

As in the other physical sciences, theories are based on the belief that the universe is homogeneous and isotropic and that events in our own galaxy are repeated elsewhere.

Nature of Information

Astronomy data, largely graphic and numeric rather than textual, retain their value indefinitely. Observers report phenomena with the highest possible degree of accuracy, with the knowledge that other researchers

will base theories on them. In fact, many observations recorded in the Middle Ages are as useful as those made yesterday.

Because astronomical research relies so heavily on observation, amateurs are also able to contribute valuable observations.

With the new instruments and computers available, there has been a giant leap in the quantity of information available, much of it now in electronic format.

The Future for Astronomy Information

According to some astronomers, most discoveries that can be made with simple instruments have already been achieved, so the scientific tools necessary to make future breakthroughs possible will have to be large, powerful, and expensive. However, others would disagree, pointing out that research using less elaborate instruments continues to yield valuable results.

If the large projects now planned succeed, the quantity of astronomy information collected will grow at such a rapid pace that its organization and dissemination will present a difficult challenge.

Use of Astronomy Information

Like other scientists, astronomers ask many different types of questions. In addition to looking for information by author or subject, they also seek information on specific objects or classes of objects. As in other disciplines, they endeavor to keep current, asking who is publishing what and what recent developments have occurred in the field. Finally, since some predictions they make are based on old information, the historical literature is essential for in-depth research.

Types of questions an astronomer might ask include, When is next year's nautical almanac going to be available? What work has recently been done on dark matter? What are the protocols for applying for research time at a particular observatory? What research has been done on a particular object?

Primary Literature

Like the other science disciplines, astronomy has seen a rapid expansion of its literature in the last several decades. It also draws on the techniques and information of a number of other disciplines, most notably engineering, physics, and mathematics, but also biology, chemistry, and the geosciences. Consequently, a certain amount of its growing body of literature does not appear in astronomy publications and may not be found in the campus library that serves astronomy.

Serial Literature

Serial literature is the most important medium for communication of research in astronomy. Astronomers say that publishers are aware of the importance of timely communication and that publication delays are minimal. It is common for researchers to browse several leading journals and to scan tables of contents of another dozen journals each month.

Older Literature

Older astronomy literature contains information that is valuable and used today. Many collections of older literature exist and are most often owned

by observatories established in the 19th century or earlier. For example, the library at the Naval Observatory in Washington, D.C., holds about 75,000 volumes covering the histories of astronomy, astrometry, and navigation and reports from other great observatories worldwide. Unfortunately, some of the older astronomy literature exists in unique copies printed on high-acid paper, which is in danger of disintegrating.

Technical Reports

Technical reports, produced by observatories and university faculty worldwide, are an important source of information, particularly of primary data. Unfortunately, many are never published elsewhere, and, probably because of their international origin, many are not indexed by the major services. Astronomers generally rely on their own devices to obtain reports they need and frequently develop extensive personal collections. At some institutions, the library also maintains a technical reports collection.

Major Indexing and Abstracting Services

Relative to the other sciences, astronomy is a small discipline, a fact that facilitates communication. Most astronomers communicate regularly with their colleagues—on the phone, at conferences, by electronic mail—and rely on informal means to become familiar with the literature. Some literature needed for research appears in the journals of related disciplines and is not always indexed in astronomy indexes. Thus it is sometimes necessary also to consult general and special indexes for other disciplines.

Print Indexes

Astronomy and Astrophysical Abstracts (*A & A Abstracts*), begun in 1969, is published by Springer Verlag in English for the Astronomische Rechen Institut, Heidelberg, under the auspices of the International Astronomical Union. Since 1990, *A & A Abstracts* has been produced in collaboration with Fachinformationszentrum Energie, Physik, Mathematik (commonly called FIZ Karlsruhe), which has resulted in an expansion of its coverage. It now abstracts some 120 journals, the publications of observatories and astronomical institutes, and significant monographs. The standard index for astronomical research, it is a continuation of the *Astronomischer Jahresbericht* (1899–1968). One drawback is that *A & A Abstracts* is only issued annually and is thus many months behind in covering the literature.

The Astronomy and Astrophysics Monthly Index, published by Olivetree Associates, began publication in 1976 and is a more timely alternative to *A & A Abstracts*. In addition to information about journal literature and observatory publications, it also includes information on preprint series from universities, laboratories, and observatories. It is not cumulated in the print version but is on a magnetic tape version. Indexing in the print publication is by title word and author.

Two print indexes are used for research in astronautics and aeronautics and related technologies. The first is *International Aerospace Abstracts* (*IAA*), a semimonthly publication of the American Institute of Aeronautics and Astronautics. In addition to aeronautics and astronautics, it covers relevant worldwide publications from other subject areas ranging from

chemistry and engineering to the social sciences. The producers will also furnish copies of papers abstracted in *IAA*.

Scientific and Technical Reports (*STAR*), a semimonthly publication of NASA, complements *IAA* by covering largely unpublished literature, including NASA, NASA contractor, and NASA grantee reports; reports issued by other U.S. government agencies and domestic and foreign research institutes, universities, and private companies; domestic and foreign dissertations and theses; and U.S. and foreign patents. Subject areas include all aspects of aeronautics and space research and development, plus the aerospace aspects of earth resources, energy development, conservation, oceanography, environmental protection, and urban transportation.

Electronic Indexes

Until recently, astronomy had no online index. But now STN has made *A & A Abstracts* available through PHYS (also known as Physics Briefs). Perhaps because this development is recent, use of online astronomy indexes is light.

DIALOG's Aerospace Database, which corresponds to *IAA* and *STAR*, provides online and CD-ROM access to published and technical report literature in aeronautics and astronautics. (The more comprehensive NASA/RECON database is only available to NASA contractors.)

Current Awareness

Conferences

Conferences, particularly specialized ones concentrating on one topic or subfield, are becoming increasingly important in astronomy. NASA, the national observatories, and subject-oriented groups within the discipline sponsor conferences where the latest research reports, many containing useful data, are presented. But conference proceedings, distributed haphazardly and outside the usual publication channels, are not widely available. Some papers eventually appear in the journal literature, but tracking a paper's transformation from conference proceeding to article is difficult. Proceedings published by major publishers are covered by the indexes, but those issued by more obscure publishers are often missed.

Newsletters

Librarians at the National Radio Astronomy Observatory (NRAO) in Green Bank, West Virginia, collect preprints reporting recent research in radio astronomy. Begun in 1978, this service is modeled on the Stanford Linear Accelerator Center weekly preprints bulletin (see **Physics**) and is equally popular among researchers. NRAO librarians compile the biweekly *RAPsheet* (Radio Astronomy Preprints) listing papers received at NRAO and send it to about 90 extramural research groups. Since many papers are never published or are only published some months or years later than the preprint, the newsletter's utility is evident.

The Space Telescope Science Institute in Baltimore, Maryland, which supports research in astrophysics and ultraviolet, optical, and near-infrared astronomy, offers a similar service, the *STEPsheet* (Space Science Preprints). The biweekly *STEPsheet* and *RAPsheet* appear alternate weeks, and the producers exchange information to eliminate overlap and ensure comprehensive coverage.

Electronic Networks

Astrophysical Journal maintains a file on BITNET of letters reporting the latest research results accepted for publication. This is a useful service, although not well known. Astronomers say that a central source of information with comprehensive coverage of current work in astronomy would be valuable.

Data Collections

To understand and identify new phenomena, theorists must have access to seemingly limitless amounts of observational data, regardless of when they were recorded. Often, later astronomers examine and explain earlier observations and reinterpret them. For instance, radio frequency radiation from space was first recorded in 1932 but not understood. It was not until 1945 that a radio telescope was built, allowing study of the radio waves and their sources and explanation of the phenomenon.

Meeting the need for data in astronomy are a number of data resources unique to the discipline.

Star Catalogs

Star catalogs have long been one of the staples of astronomy research. Old bright star catalogs are still as useful as when they were first produced, since, as one astronomer pointed out, "the bright stars are still in the same place." The *Henry Draper Catalog*, first published in 1918, is considered a standard. But catalogs of these older observations are poorly controlled. Those produced before 1970, when *A & A Abstracts* began to include them, are not listed in a comprehensive index or bibliography.

Astronomical objects are known by their designation, that is, a name and number taken from the catalog in which they are noted. Objects from the *Henry Draper Catalog* and its extensions are known by their HD number. Some galaxies and star clusters are known by their NGC number, taken from the *New General Catalog of Nebulae and Clusters of Stars*. However, there are literally thousands of catalogs, and one object may have as many as 20 designations because of an observer's ignorance of a previous observation or possibly a desire to have his or her name attached to it. HD 1486, for example, has also been listed in *Wilson's General Catalog of Stellar Radial Velocities* (1953) as number 175, thereby giving it the designation of W175. The designation W175 is also given to other objects in the catalogs of Wolf (1913), Wooley (1970), and Wray (1966). Comprehensive collections of star catalogs enable astronomers to sort through the welter of designations, but few libraries have historical collections strong enough to support in-depth research. Eventually, some 500 catalogs will be collected and cross-indexed in a database being developed at NASA's Astronomical Data Center (ADC). Thousands more will remain to be converted.

Photographic Plates

Photographic plates are an important resource, especially for observations of single-occurrence events, but many are quite old and deteriorating rapidly. They are expensive to reproduce, but fortunately it is now possible to put them in electronic format, which will preserve the intellectual content if not the plate itself. In fact, the photographic plates of several big surveys such as the Palomar Sky Survey have recently been digitized

and will be available to astronomers in the near future. Although the magnitude is still too "coarse" for comparing any but the brighter stars, the electronic format allows for correlating them with objects that emit radio waves.

Other Electronic Resources

Vast amounts of astronomy data that were once available only in print or on photographic plates are now available in electronic format. This change has made it possible to collect, correlate, and compare data in ways never before feasible. With powerful computers, large databases can be constructed, allowing entire populations of astronomical objects to be studied. Also with the aid of computers, it is possible to mix data from different wavelengths and correlate observations of x-ray, infrared, ultraviolet, radio, and optical data.

Space Physics Analysis Network

NASA, the Department of Defense, and other agencies have over the past two decades supported packet-switching networks for sharing research and technology. SPAN (Space Physics Analysis Network) began operating in 1981, connecting more than a thousand networks. Using these networks, scientists gain access to data repositories and computing resources worldwide. They are also able to communicate with colleagues in academia, government, and industry in this country and abroad.

NASA Data Sets

It has been estimated that NASA has created more than 1.2 million magnetic tapes since the agency's creation in 1958; many of these contain information of great potential value to astronomers. Yet much of the information gathered has been so poorly labeled and stored that it is all but—or even entirely—inaccessible. The problems are familiar: code books describing how to use the tapes have been lost, tapes have been physically damaged, and the hardware necessary to read older tapes no longer functions. Most significant, no standard method for describing and cataloging the tapes exists.

NASA is making an effort to improve control over its vast store of information. At the Jet Propulsion Laboratory, Pasadena, California, 135,000 tapes from two dozen space missions are being restored; some of the data are also being distributed on CD-ROM through the National Space Science Data Center. NASA is also developing a central directory of data generated both by NASA and by other sources (see **Geosciences**). Nonetheless, astronomers say that important data collected by satellites and by telescopes at the national observatories are still difficult to identify and track down. Until intellectual and physical control over NASA data is given as high a priority as data collection, this situation is unlikely to change.

SIMBAD

Telescopes in observatories worldwide collect data every night from millions of stars and stellar systems. Many of these data are brought together in the most complete astronomical database in the world, SIMBAD (Set of Identifications, Measurements, and Bibliography for Astronomical Data), which is maintained at the Centre de Données Stellaires in Strasbourg, France. It allows researchers to look up an astronomical object such as a star or supernova by its astronomical designation and to get both information about the object and a list of papers published about it since

1950. SIMBAD includes observations collected and analysed by NASA and astronomical institutes across Europe, and became available in the U.S. in 1989 through SPAN.

Satellite Imagery

Using laser infrared photography, satellites have photographed vast expanses of the sky. This information has been digitized and is available in electronic format, generally on magnetic tape. In the case of the Hubble Space Telescope, data collected will be stored in an optical system, the Data Archive and Distribution Service. Twelve-inch WORM (Write Once Read Many) disks, mounted in "jukeboxes" holding 100 or more disks, will store the data. The system is projected to have a capacity of 51 terabytes. The Planetary Data System at the Jet Propulsion Laboratory has also begun distributing data on optical disk.

Private Databases

Observational astronomers have produced or compiled databases for use by their own research groups or by affiliates. Many astronomers have said that they would be pleased to share this data with others, and some data sets have indeed been made available to the research community. However, the majority of data sets remain inacessible or virtuallly so. At the moment, several obstacles impede data sharing. First, some express concern because home-grown databases are often not constructed systematically and may contain errors which, if not eliminated, could lead to errors in interpretation of real or extrapolated measurements. Second, there is no central listing of data sets. Personal connections are the principal means of locating them, but no systematic method of determining what is available exists. Third, data for each wavelength are collected according to different standards and are usually archived separately. Some data sets will be made more accessible through a project at NASA's ADC to acquire, edit, and modify them, but many more are likely to remain out of circulation.

Future Needs and Directions

Astronomy Literature

Even with more electronic star catalogs under development, pre-1970 star catalogs will remain essential to the discipline, whether used alone or in combination with other catalogs to extract a variety of types of data. Better bibliographic control of these older catalogs is needed, ideally by making records available on the national bibliographic utilities and in the online astronomy indexes.

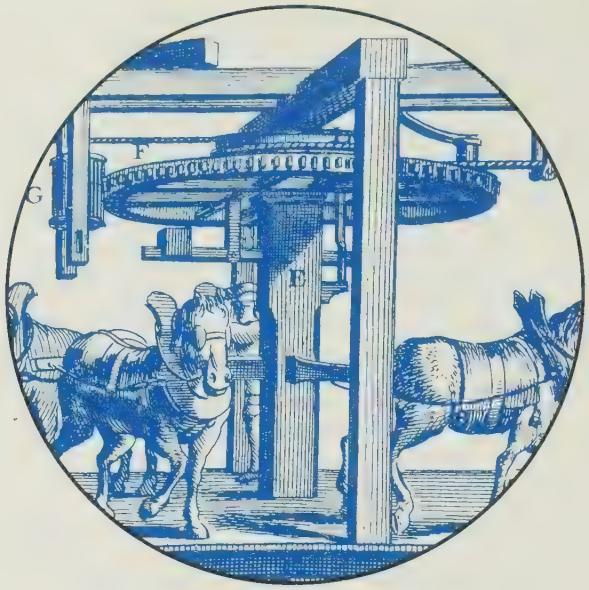
Bibliographic access to unpublished conference literature should be improved and made available on the national bibliographic utilities. In general, a central source of information about recently published research and research in progress in subfields that have no such resource would be useful.

Electronic Resources

Research in astronomy will depend to a growing degree on electronic data sets and on intellectual and physical control over them. A central listing of data sets, possibly developed cooperatively by research libraries and the national observatories, is needed. A strategy to include information on private data sets should also be developed.

Preservation

Projects to preserve deteriorating photographic plate and print collections, essential resources for astronomy research, are needed.



Engineering

General Characteristics

Nature of Research

Engineering applies the principles of physics, chemistry, and mathematics to the solution of practical problems and the design of devices.

Engineering has traditionally been grouped into the subdisciplines of chemical, civil, electrical and electronic, industrial, mechanical, and mining and metallurgical engineering. With the emergence of newer subfields (such as aerospace, biological, materials, environmental, and nuclear engineering) that overlap with other science, social science, and even humanities disciplines, the scope of engineering has become enormous.

Engineers are the ultimate problem solvers: their goal is to design things that are useful.

Most academic engineers work in small- to medium-size groups composed of several principal investigators assisted by research associates and students. They want access to information in their office or laboratory rather than in the library.

Industrial ties with academia are strong. Engineers in industry, particularly in research and development groups, may work closely with academic engineers.

Computers have revolutionized engineering. Slide rules and drawing boards, the engineer's traditional tools, have virtually disappeared. Huge matrices of numerical formulae can now be analyzed; materials and designs can be modeled and tested on the computer before they are actually manufactured. Property data that once had to be found by searching dozens of handbooks is beginning to be available in online databases.

Nature of Information

Engineering information, reflecting the enormous breadth and diversity of the field, encompasses data from all the sciences. Data sought may have a wide or very narrow application, depending on the engineering project. This contrasts with chemical information, which focuses on molecular structure and is widely useful to many types of chemistry research.

Older data are often useful. When petroleum became more expensive in the mid-1970s, for example, research in other fossil fuels going as far back as the 1920s was heavily consulted.

The Future for Engineering Information

Increasingly, traditional information will be dispensed in nontraditional ways (e.g., on CD-ROM and electronic networks), and the demand for electronic information is likely to grow. To allow faculty to take advantage of innovations in information delivery, the information must be accessible in offices and laboratories.

Academic engineering research is critical to virtually all types of industry, which also invest heavily in basic research. To remain competitive, industry must have access to engineering technology and information.

The demand for access to engineering research from Japan, Western Europe, and other industrialized countries outside the U.S. will grow.

In the past 20 years, journals have proliferated and the amount of engineering information has grown rapidly. This trend also seems likely to continue.

Use of Engineering Information

Above all, engineers are problem solvers. They are also concerned with predictability—to design a device or system on paper so that when it is built it will function as predicted. The questions they ask are generally focused on understanding what is happening in a given system, whether mechanical or subatomic. This requires knowledge of general scientific principles, mainly drawn from physics, mathematics, and chemistry, coupled with experimental data. After general analysis of the problem, they will ask specific questions of the engineering literature, such as: What is the thermodynamic property of a certain metal under particular conditions? What is the effect of one compound on another? How should a particular type of tube be coated? How should one test for a leak? What is the standard for a particular programming language or for testing the compressive strength of cement?

The information sought is so varied that researchers and librarians grope their way among a variety of information resources, with varying degrees of success.

Primary Literature

Engineers use virtually all forms of published material. Manuals and handbooks for project design, books for background material, journals and symposia for both current and older information, computerized databases and indexes—all are widely used and valuable sources of information.

Serial Literature

Serials are important more as a permanent repository of information than as a source of the latest information. Nonetheless, most engineers read or browse a set of core titles regularly. Like other scientists, they complain about the proliferation of journals and about the length of time an article takes to appear. They also comment that articles are often shorter than formerly and tend to contain fewer data. The greatest advantage to journals as a source of information is that articles have usually been reviewed.

Professional engineering societies have long provided the core of engineering literature. Since virtually every subfield of engineering has its own professional society, the list of engineering societies and their publications is long. With the number of researchers increasing and specialization growing, journals that cater to subfields and specializations within subfields have proliferated in the past 10 to 20 years. At the same time, subscription costs have increased much faster than the rate of inflation, threatening continued access to serial literature. Libraries hope to develop cooperative serials projects that will allow them to continue to provide access to the range of information their researchers need while keeping costs under control.

Patent Literature

Patent literature has long been important to corporate engineers, and is becoming increasingly important in academia as well. Particularly for corporate researchers, the patent literature is used as a means of tracking the activities of competitors, as well as protecting inventions. For several reasons—including the higher profile of university technology licensing offices, the wider reach of patents (e.g., to include software), and increased collaboration with industry—academic engineers are more likely now to have a need for patent literature than in the past. (For a discussion of patent literature indexes, see **Chemistry**.)

Technical Reports

Technical reports are a mainstay of engineering research in many subfields, at least in part because they contain substantially more data than journal articles and are more current sources of information on the progress or results of research and development. Issued by a welter of individuals, departments, and laboratories, they can be elusive. Distribution is limited and somewhat haphazard; many laboratories maintain their own collection of reports, which get thumbed and worn and sometimes lost. Rather than turning to indexes to locate technical reports, established researchers tend to rely on their personal networks to yield up what they need. Although technical reports are not refereed, engineers are willing to consult them to get the information they need.

For researchers who do not have established networks or who seek information outside their immediate field, the National Technical Information Service (NTIS) provides access to bibliographic information on research sponsored by the U.S. and other governments. Many reports of interest to engineers can be found in the NTIS database or its print counterpart, *Government Reports Announcements and Index*. NTIS also provides copies. Other sources include the print indexes *Scientific and Technical Aerospace Reports* (*STAR*), available online in DIALOG's Aerospace Database and the DOE Energy database (see **Astronomy**). However, they

often miss research that is not government funded, including reports from academic institutions and industry.

Not found in the above indexes are technical reports distributed by companies, sometimes in the form of scientific marketing literature. Generally, companies do not maintain a file of their own reports, which can contain valuable scientific tables or data about the performance of scientific instruments. Once the supply has been exhausted, it may be difficult to trace the report later.

A technical report may contain the germ of an idea 30 years old that can be explained with a new theory. It may describe a manufacturing process that has not changed substantially in 20 years. It may establish scientific precedence. Consequently, older technical reports are often called for but require more ingenuity to track down than recent ones.

Standards Literature

Engineers of all types use standards information. The standards literature includes specifications, codes of practice, guidelines, nomenclature, and terminology rules for products ranging from nuts and bolts to computer programming languages. Development of standards is highly decentralized, and thus the sources of standards are numerous: companies, technical associations, professional societies, government agencies, and national and international standards bodies. To bring some coherence to this arena, the National Bureau of Standards (now the National Institute of Standards and Technology) was established in 1965. One of its principal functions is to collect standards in active use in the U.S.

Information vendors such as Information Handling Services index standards in print, online, and on CD-ROM; they also supply sets of standards with product specifications on microfilm for almost all U.S. manufacturers. Few libraries collect standards information, as it is expensive and updated frequently, but order standards as needed instead. However, despite the existence of organizations devoted to minimizing the fragmentation of standards information, librarians report that it is often necessary to contact a number of agencies or people to track down a particular standard.

Major Indexing and Abstracting Services

Established researchers say that, before consulting indexes or standard information sources to locate serial literature, they prefer to rely on their personal collections and on information supplied by colleagues. Graduate students are more apt than established researchers to use bibliographic databases to identify information sources and thus more likely to turn up something new in an obscure journal.

Print Indexes

Of the general indexes devoted to engineering literature, *Engineering Index* probably serves the broadest audience (although other indexes may serve particular specialties better). Begun as an annual in 1884, it became a monthly in 1906. It covers the world's technical literature in all engineering disciplines, offering full bibliographic information and abstracts of journal literature, technical reports, monographs, and conference proceedings. A little more than half of the publications covered are

from Western Europe, Canada, the Soviet Union, and Japan. All abstracts are in English.

Science Abstracts, produced by the Institution of Electrical Engineers, was first published in 1898. In 1903 it was divided into two sections: Section A, *Physics* (see **Physics**), and Section B, *Electrical Engineering*. In 1966, Section C, *Computer Control Abstracts*, was added, and Section B became *Electrical and Electronics Abstracts*. *Electrical and Electronics Abstracts* aims to cover the world's electro-technology literature comprehensively. Published monthly, it adds about 84,000 items a year. Its abstracts are also in English.

Science Citation Index, a multidisciplinary index (see **Biology** and **Physics**) was started in 1961. Although it covers a smaller number of journals than the two above indexes, engineers find it useful for its unique approach to indexing.

Engineers also use a multitude of more specialized indexes. Some of these include *Energy Research Abstracts*, *Selected Water Resources Abstracts*, *Metals Abstracts*, and *Engineered Materials Abstracts*.

Electronic Indexes

Engineering Index is available online as Compendex for most articles published since 1969. Ei Engineering Meetings, which contains abstracts from papers presented at conferences abstracted in *Engineering Index*, was merged with Compendex in 1988 into a single integrated file, called Compendex*Plus, containing abstracts of both journal articles and papers since 1982. (Some vendors still offer the two files separately.) The combined database contains more than 2.5 million records to the worldwide literature, excluding patents, in engineering and technology. A major advantage of the database is the inclusion of treatment codes, which enable the user to search for applications, new developments, and theoretical information, as well as practical or experimental information. Compendex*Plus is also available on CD-ROM.

INSPEC, the online version of *Science Abstracts*, Sections A, B, and C, contains citations and abstracts to more than 3.3 million articles in computer science, electrical engineering, electronics, and physics published since 1970. Like Compendex*Plus, INSPEC also allows searching by treatment codes.

Science Citation Index is available online as SCISEARCH and on CD-ROM.

The Japanese are doing important work in all areas of technology but few research reports are available in English. The Japan Information Center of Science and Technology (JICST) offers abstracts of science and technology literature, including Japanese journals, technical reports, and conference proceedings published since 1985 in STN's JICST-E database. Although citations and abstracts are in English, the text is in Japanese and translations are only available on demand, at relatively high prices.

Librarians report that translations take a long time to obtain and that some publications of interest are not included in the database.

For information on current research, STN offers the Japanese Government and Public Research in Progress (JGRIP) database, which contains information collected by an annual survey of public research institutes in Japan. Information includes research topics, research institute names, researcher names, and project start and completion dates. Other resources providing information about Japanese publications are available, but librarians report that access to the information they need is not always adequate.

Several databases provide access to the standards literature. Industry Standards and Military Specifications, produced by Information Handling Services, contains more than 200,000 citations to government and industry standards, including those of 70 national and international standards organizations. Some information dates back to 1895. Standards and Specifications, produced by the National Standards Association, Inc., contains citations to more than 100,000 U.S. and international standards documents.

Numerous other online indexes also provide coverage of the literature, both of specific subfields (for example, Chemical Engineering Abstracts) and of related and interdisciplinary fields (for instance, DOE Energy and CA).

Because of their quantitative orientation, many engineers tend to think of computers as tools for analysis but not as sources of bibliographic or factual information. Many are unfamiliar with the databases available in their field, in part because their libraries or information centers have not educated them to do online database searching.

Nevertheless, a significant number of engineers use the online services offered by librarians and information scientists in engineering libraries and information centers, and, as more resources become available, especially on CD-ROM, use seems to increase. Although some engineers search online databases unassisted, most prefer to have searches performed by search intermediaries. The protocols are very detailed and the charges substantial enough to discourage infrequent use.

Current Awareness

Current Research

Like other scientists, engineers depend on information about research others are undertaking. One source of this information is the Federal Research in Progress database, produced by NTIS, which provides information about current research at the U.S. Departments of Agriculture and Energy, NASA, the National Institute of Standards and Technology, the National Institutes of Health, the U.S. Geological Survey, the National Science Foundation, and several other large agencies. Even this wide coverage is not comprehensive, however, since it excludes some research done at private and university laboratories.

An earlier project, the Smithsonian Scientific Information Exchange, available through DIALOG as SSIE Current Research, provided broader coverage, including research sponsored by state and local governments, major foundations, individuals, and universities. Ironically, although it was conceived as a way to track current research but only covers the period 1978 to 1982, it is used even now because the information is still valuable.

Conference Proceedings

Increasingly, engineers depend on conference proceedings to communicate research results. Even though proceedings are not refereed, they are more up to date than the journal literature and often contain more data. Particularly useful are papers from specialized conferences focusing on one area of interest within an engineering field (e.g., heat transfer or signal processing). Although the papers have not been reviewed and it is difficult to determine what is of lasting value, the advantage of quick access outweighs the disadvantages. Proceedings from these conferences are often only available in a limited number and generally are not published, although papers may appear later in altered form. They are also not collected systematically and are not indexed. Consequently, researchers must rely on their own devices to obtain them. Papers from meetings sponsored by one of the major engineering societies are more likely to be published and indexed.

Engineers in industry use conference proceedings both to obtain information about related research and to keep an eye on their competitors. In Europe and Japan, quick-issue patents serve this function, but the U.S. patent system, which releases information only after the entire patent process is complete, is at least several years behind and therefore does not serve a current awareness function.

Serial Literature

Engineers have several means of staying current with journal literature. Many engineering libraries photocopy and circulate tables of contents of journals of interest to their researchers. For more comprehensive coverage, *Current Contents*, a weekly publication from the Institute for Scientific Information (ISI), carries contents pages from core journals in broad scientific fields. In addition to the tables of contents, it offers a complete bibliographic record for each item plus journal information and authors' addresses where available; it is also available online. The *Current Contents* section "Engineering, Technology and Applied Sciences" covers many journals of interest to engineers.

Some database vendors offer what are called selective dissemination of information (SDI) services to researchers. Using a profile of a researcher's interests, the vendor performs searches on one or more databases at regular intervals and sends the results to the researcher. In many cases, librarians also provide this service locally.

Finally, current awareness services in specific subject areas, such as those offered by the *Journal of Polymer Contents* and *Current Literature in Traffic and Transportation*, are useful for researchers in the relevant fields.

JICST offers a specifically tailored current awareness service to Japanese technical journals with abstracts in English. Subjects covered include biotechnology, materials, electronics and computer technology, medical technology, and robotics.

Other Electronic Resources

Some engineers have little need for factual, nonbibliographic databases. In many cases, too much explanatory material is necessary to make quantitative data generated by another engineer's project broadly useful, so computers are employed mainly for analysis rather than storage and retrieval of data. Generalized data eventually turn up in papers but have not until recently been brought together in one database intended for many purposes. Unfortunately many data are often lost or simply not reported. With skyrocketing page-production costs authors are discouraged from publishing supportive data. They often will publish "pseudo-data" that are greatly reduced or simplified and cannot be verified without great effort. The formats in which the data are presented may not be useful to others.

Not surprisingly, database producers are beginning to respond to the need and potential for databases which contain quantitative information and allow for searching by property for a bibliographic reference. Several projects now available or in progress will enable engineers to consult large data files of physical, electronic, and materials property information that have been gleaned from articles and texts by identifying the parameters of the data.

Probably the largest of these is the Materials Property Data Network, scheduled for release by STN in 1991. It will consolidate a number of materials property databases, which must now be consulted one at a time, to include a broad and heterogeneous array of data on the mechanical, electrical, thermal, magnetic, thermophysical, electrochemical, and corrosion properties of metals, polymers, ceramics, and composite materials. INSPEC offers data on semiconductors, silicides, and other materials that are used in semiconductors through EMIS, the Electronic Materials Information Service. Beilstein Online (see **Chemistry**) offers information on organic compounds.

The growing list of data needed includes design codes and adsorption data. However, enormous effort and expense are required to document and standardize this type of information. Further, because the data are very complex they require large amounts of computer memory to store. Engineers point out that it may take 75 pieces of information to describe one material, hundreds more to describe its properties, and up to 300 to describe one test.

Future Needs and Directions

Engineering Literature

Quick access to all types of literature, both current and retrospective, will continue to be a fundamental need. Several factors will affect access:

- Academic libraries will remain the chief sources of research information. However, the rapidly increasing cost of scientific and technical literature may force many libraries to reduce subscriptions and to rely

on interlibrary loan and commercial document delivery services. Thus, the need for efficient and inexpensive document delivery systems for article-level information is likely to grow.

- Access to bibliographic information on research not included in the major indexes—grants and certain conference proceedings and technical reports—needs to be improved. This will facilitate communication of research results within the discipline and access to data omitted from journal publications.

Integrated Information Environment

As increased amounts of information become available in electronic form, engineers will seek a more integrated information environment. The configuration of this information environment will include access to national and international computer networks; local area networks containing the full text, with graphics, of the latest journal literature; important handbooks; materials property databases; gateways to commercial databases; and software collections. Cost will be an important factor, especially for academic libraries and researchers.

Education

Researchers and librarians alike remark that, with so many information products already available and new ones continually cropping up, the field is prey to information confusion. This situation presents a challenge to the librarian, who is probably best qualified to link the user with the information sought. Especially as electronic information resources increase and libraries shift their emphasis from ownership of materials to access, graduate students and faculty will need training in their use.

Data

The growth of information storage and retrieval capabilities opens up opportunities for sharing evaluated physical property and other data. Engineers and engineering societies can take the lead in promoting the development of databases containing basic data for research.

Mathematics

General Characteristics

Nature of Research

Mathematics deals with magnitude, quantities, numbers and their relations, and the properties of space. It interacts closely with the other sciences, especially the physical sciences. On the one hand, it is stimulated by problems in the physical sciences; on the other, as the reasoning tool for the sciences, it contributes to the solution of their problems.

Historically, mathematics has been the product of human thought, unaided by technology or laboratory equipment. For this reason, mathematicians in less developed countries are able to make important contributions to the discipline, which is less frequently the case in other fields of science.

However, mathematicians increasingly use computers for a variety of tasks. Formulation and testing of hypotheses are now possible for complex problems that were once out of reach. They also use computers for literature retrieval, communication via networks, and, to a small extent, scholarly publishing.

Some mathematicians regard their goal as producing a work of art. Utility is not avoided, but it is considered inferior to what has been called an "intrinsic elegance and profundity." Nonetheless, applied mathematics—statistics, computing, numerical analysis—remains important and is often difficult to distinguish from "pure" mathematics.

Nature of Information

Knowledge in mathematics is cumulative, and the different areas—e.g., algebra, geometry, calculus—are closely intertwined. Mathematical literature retains its value over a long period of time, and mathematicians frequently make use of the core literature.

Although mathematics is thousands of years old, the body of mathematical monographs and journal literature is small relative to that of other science disciplines.

In addition to access to older literature, mathematicians also need quick access to recent literature.

TeX, a software program that allows standard printers to print mathematical symbols, has been a boon to the discipline because it facilitates both printing and information retrieval with mathematical symbols.

The Future for Mathematics Information

Electronic mail and telefacsimile have replaced surface mail and the telephone for some mathematicians, and will probably continue to make inroads into the discipline. Nonetheless, it seems likely that the methods of information use and dissemination in mathematics will change at a slower pace than in other science disciplines.

Mathematicians will continue to benefit from advances in information technology such as online bibliographic databases and tools like TeX. But because the development of technology is uneven worldwide, existing conventional publication methods will probably hold sway over newer methods for some time to come.

Use of Mathematics Information

Questions in mathematics run from the very general (e.g., Where can I find something about . . . ?) to the more specific (e.g., Is there a version of this theorem that is valid when the third hypothesis is not quite the same as in this paper?). General questions in particular are often addressed to a colleague.

Often a mathematician asks whether a result he/she has established or conjectured, or would like to apply, has been published. He/she might want to determine if a proof would be true with a weaker hypothesis. The person might be looking for links within other mathematics fields to what he/she is working on. To answer the question, the mathematician consults a number of sources until an abstract of a paper that appears to be relevant turns up. At that point, closer inspection of the literature is in order. For that purpose, mathematicians prefer to have their own journal volume or photocopy to examine in their office. The goal is to avoid spending time working on something that has already been done.

As in other disciplines, mathematicians also at times seek to gain familiarity with a body of new ideas as presented in a monograph or major research article. Again, they prefer to consult a printed copy at home or in their office.

For browsing, *Mathematical Reviews* (see below), containing both abstracts and reviews of the literature, is often helpful, particularly in the print or CD-ROM version.

Primary Literature

Information needs in mathematics are probably the most straightforward of any science discipline. Beyond pencil and paper, the mathematician primarily requires access to current and historical literature. The data required by other disciplines—genetic sequences, chemical compounds, or other data based on observation of natural phenomena—are irrelevant to mathematics. Electronic and other nonprint media are little used to distribute published literature although they are beginning to be used more frequently for the dissemination of preprints. Nearly all significant published information is in print format.

Serials and Monographs

Serials and monographs, new and old, are absolutely vital to the research of most mathematicians. Active use of old serials, some dating from the early 19th century, is common. Use of older monographs is also frequent. Several of the leading mathematics collections in the U.S. have been cataloged comprehensively in the national bibliographic utilities, making them accessible to the research community. However, librarians point out that one or two locations for a monograph or serial do not adequately support research use and that the addition of information about other important mathematics collections would enhance research access.

In the case of older materials, preservation issues affect access. At some institutions, monographic series and serials have been microfilmed for preservation purposes. Users, however, object to the inconvenience of microfilm, especially for monographs. It is more suitable for serials, since a single article can be printed without great expense.

Most mathematicians prefer a departmental library devoted only to mathematics because it makes browsing easy. They like to know they will find the green book halfway down aisle three that they have consulted before. Some also like to have access to a wide range of materials on mathematical theory in computer science, mathematical physics, and other areas in addition to core mathematics information.

Preprints

As in other disciplines, communication with other researchers is vital to mathematics research. In mathematics, preprints are the most important medium for consultation among scholars, for sharing information, and for avoiding duplication. However, the circulation list is sometimes a constraint to information sharing. It is usually developed by the author and does not necessarily include a comprehensive list of colleagues who are doing related research. Thus, many of the latter, particularly those at less prestigious institutions, may not learn of a preprint's existence for some time. This is unfortunate, since mathematicians say that in some subfields an article may be out of date by the time it appears in print.

Translations

Although English is the prevailing language spoken at international conferences, a significant number of publications necessary for research in mathematics come from the Soviet Union, Japan, Europe, and developing countries and are not written in English. Because it is important for mathematicians to keep up with international developments, a good mathematics library ensures that the international literature is represented, mainly by purchasing the leading mathematics journals from a number of countries.

Many important foreign-language mathematics journals are translated cover to cover by commercial and scholarly society publishers. However, many academic libraries find them prohibitively expensive. Over half a million translations of scientific articles are available through the National Translations Center at the Library of Congress. (Until January 1989 it was located at the University of Chicago's John Crerar Library.) Because the contributors of translations are predominantly corporate libraries, the center's collection is slanted toward research in applied sciences rather than theoretical, rendering it more useful for corporate researchers than academic.

Major Indexing and Abstracting Services

Mathematicians are fortunate that mathematics indexing tools do not emphasize serial literature at the expense of monographs. Both are covered by indexing and abstracting tools, and most mathematicians use at least one of the mathematical subject indexes. Their usefulness is enhanced by the inclusion of literature related to mathematics from fields such as statistics and computer science.

Print Indexes

Mathematical Reviews (also known as *Math Reviews*), published by the American Mathematical Society (AMS), provides abstracts and reviews of the main research publications of mathematics, as well as statistics, computer science, and mathematics-related work in other disciplines from 1940 to the present. In addition to selected monographs, 1,600 journals are covered in *Math Reviews*, and 500 are abstracted in their entirety.

Zentralblatt für Mathematik/Mathematics Abstracts (usually referred to as *Zentralblatt*), which began publication in 1931, is another comprehensive mathematics index. Like *Math Reviews*, it goes beyond pure mathematics to cover computer science, statistics, and other mathematics-related topics. Although it is produced in Germany, all citations are in English; it is available in the U.S. but not heavily used. An advantage to *Zentralblatt* is a relatively short lag time before it appears because its abstracts are submitted by the author along with the article rather than by an independent reviewer.

Electronic Indexes

Math Reviews is available online as MathSci, an online database with bibliographic data from articles published since 1959 and the entire text of reviews since mid-1979. The database also includes *Current Mathematical Publications* (*CMP*), a brief list of all new items received by *Math Reviews*; *Current Index to Statistics*, a comprehensive index of publications having to do with statistics in a broad range of disciplines; *Index to Statistics and Probability*, which provides retrospective coverage of the literature of statistics published between 1902 and 1968; and two publications of the Association of Computing Machinery (ACM), *Computing Reviews* and *ACM Guide to Computing Literature*. Mathdoc, a fee-based document delivery service, provides copies of articles cited in MathSci.

The current ten years of *Math Reviews* and over 50,000 entries from *CMP* are available quarterly on CD-ROM as MathSci Disc. Although many mathematicians welcome the flexibility and freedom from per-search charges

that CD-ROM provides, some predict that MathSci Disc's quarterly publication schedule may not prove adequate to meet the need for timely information. The limited time-span coverage is an added drawback; according to mathematicians, anything less than the full database is of limited use.

Zentralblatt is available online as MATH, with citations in English to worldwide publications in pure and applied mathematics issued since 1972.

Mathematical research is dependent on the use of symbols not found on standard keyboards. Because of the limitations of most printers and typesetting devices, online retrieval of titles from MathSci, which often contain non-Roman symbols, is very cumbersome. The TeX software program (mentioned above) has become standard for reproducing mathematical titles. But titles and abstracts entered in databases before TeX became available (1985 in the case of MathSci) can be difficult to decipher because they do not have the benefit of TeX coding.

Current Awareness

The above discussion of preprints makes it clear that access to unpublished research through the preprint network is vital to current awareness in mathematics.

Mathematicians also require current access to published literature, and most take time regularly to browse through current journals. For those who need more systematic access to current literature than browsing provides, the indexes are essential, particularly if they make information available in a timely manner. To bridge the gap between the publication of articles, books, and conference proceedings and the appearance of the review, the AMS publishes *Current Mathematical Publications* (*CMP*) every three weeks. As mentioned, *CMP* lists all new items received by *Math Reviews*; it also contains page proof copies of tables of contents of important mathematical journals and *Math Review*'s subject listing of articles by author and title. Primary and secondary classification numbers are assigned to each entry, which also includes the authors' institutional address.

Future Needs and Directions

Serials and Monographs

Like scientists in other disciplines, mathematicians are worried about the rapidly rising cost of journals. This problem, coupled with space constraints that many libraries are experiencing, will bring about reassessment of acquisition strategies in a number of libraries. Some predict that libraries will have to focus on literature serving the research programs of the department, the most significant international literature, and preservation of older materials. Purchases of subject-specific journals, particularly those outside the purview of the department's immediate interests, may have to be cut.

Cooperative library programs can help to ensure that a broad spectrum of research materials is still collected in the U.S. and is available to researchers. Through RLG's Long-Term Serials Project, RLG libraries have developed a method of sharing long-term commitments to subscribe to a

wide range of mathematics journals. This project, and others like it, will be increasingly important in the future.

A corollary of the collection situation is that, increasingly, mathematicians will be required to go beyond the department library and to rely on libraries at other institutions for the information they need. An efficient and affordable document transmission capability will be necessary to support their needs.

Preprints

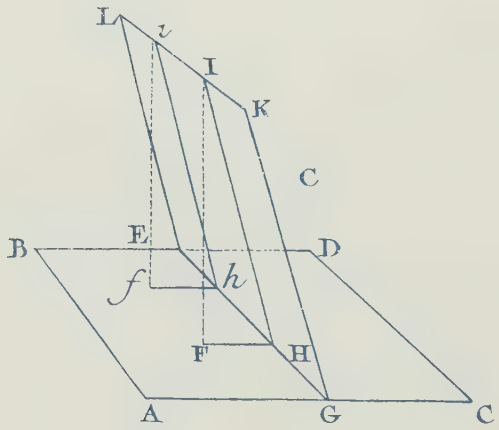
Access to preprints, which is necessary to be informed about the very latest research, is often problematic. Several mathematicians have suggested that an electronic network for distributing preprints or abstracts in mathematics would be useful. Indeed, the Instant Math Preprints Project, a pilot project led by Yale University, will make abstracts and the full text of math preprints from several universities available. If successful, the project could expand to include other mathematics departments in the U.S.

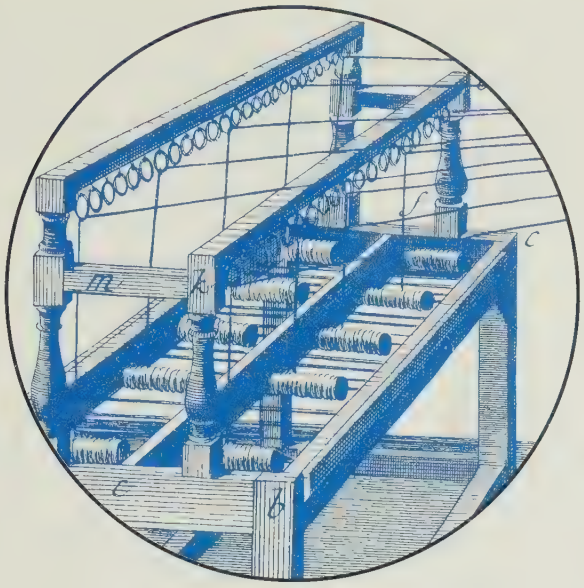
Preservation

Hard copy reformatting (through photocopy) of older monographs is the preferred way to provide access in many libraries. Since photocopying is not coordinated among libraries, this practice has led to duplication of effort. This is generally not the case with preservation microforms, which are produced only after a careful search for records of existing master microforms is made. Cooperative preservation projects, in conjunction with retrospective conversion of card catalog records of important older collections, would improve access to older materials.

Electronic Publishing

Because the mathematics community is international in character and many mathematicians outside the U.S. have little or no access to modern technology, it may be a long time before significant changes in publication practices take place. Existing, conventional methods may serve the international mathematics community best for some years to come.





Computer Science

General Characteristics

Nature of Research

Computer science studies the theory, analysis, design, and application of the algorithmic processes that describe and transform information. The fundamental question computer science poses is, what can be efficiently automated and how can it be done?

It is a discipline broad in scope and heterogeneous in nature. Within the discipline, there is no general agreement as to its boundaries and purpose. To some, it is closest to mathematics; to others, engineering. Some emphasize its connections with philosophy and logic.

Computer science can be divided into a number of subfields, including algorithms and data structures, programming languages, architecture, numerical and symbolic computation, information retrieval, operating systems and networks, software methodology, artificial intelligence, and visualization and robotics.

Because of the nature of computer science, it is served by three professional societies: the Association for Computing Machinery (ACM), which tends to be favored by those who are more mathematically or theoretically oriented, and the Institute for Electrical and Electronics Engineers (IEEE) and the Society for Industrial and Applied Mechanics (SIAM), which attract those with an engineering or applications orientation.

Computer scientists work in well-equipped computing environments. Yet most computing in the discipline is on a relatively modest scale, not requiring supercomputers or large data sets.

Nature of Information

Much computer science literature is of recent vintage, reflecting the youth of the discipline, which began to take shape after World War II. More than in other disciplines, a large portion of it quickly goes out of date, although some literature used in computer science research (e.g., mathematics literature) remains useful almost indefinitely.

Computer scientists regard wide area networking as essential to sharing information and supporting collaboration between researchers at different sites.

The Future for Computer Science Information

Networking, which has come into widespread use in the last 15 to 20 years, has profoundly changed the face of computer science and will continue to shape its course.

The velocity of change in the field is likely to increase. Amidst this rapid change, debates about the discipline's identity will continue and computer science will continue to seek to define itself.

Use of Computer Science Information

As in other fields, questions in computer science vary widely. They might range from a comprehensive subject search in an index to a more general quest for recent developments in a particular field. Some computer scientists browse the literature in search of ideas. Examples of questions asked in computer science include: For given classes of problems, what are the best algorithms? What are the basic construction rules for large systems? What are the protocols for short- and long-haul communications? What are fast methods for selecting and joining elements from large sets of data?

Primary Literature

Serial and Monographic Literature

Journal literature, although important, is less central to computer science than it is to other science disciplines. The body of literature is relatively small and mainly of recent vintage. The oldest technical reports date to the early 1950s, and ACM's *Computing Reviews*, one of the field's early indexes, only began publication in 1960. Because the discipline and the technology on which it depends are changing so rapidly, computer scientists use retrospective literature infrequently. According to one researcher, most computer science literature is obsolete after five years. Thus, computer scientists are on the whole less dependent on access to the historical archives of the research library than other scientists.

But like any group of creative thinkers, computer scientists resist such easy generalizations. One computer scientist, for example, enjoys using the library to trace the mathematical origins of the discipline in Arabic mathematics. As mathematical information does not go out of date, access to old and recent mathematics serials and monographs is crucial for some computer scientists.

Although English is the predominant language of publication, some computer scientists have occasion to use publications in languages less accessible to English speakers—mainly Japanese, Chinese, and Russian. However, it is often difficult to obtain translations in a timely fashion.

Technical Reports

Computer scientists tend to rely more on technical reports from large and small university departments and from industrial research and development laboratories such as IBM and AT&T than on traditional published literature. Researchers are primarily interested in recent reports, although older ones, particularly those with an emphasis on theory, are sometimes requested. Established researchers generally obtain recent reports directly from the author, but technical report collections in libraries are important to less well-connected researchers, and for older material. Many computer science departments also have exchange programs with other departments and laboratories. An enormously valuable resource is Stanford University's listing of mathematics and computer science technical reports, considered one of the most comprehensive of its kind.

Conference Proceedings

In computer science, conferences are the venue for presenting important new research, and competition for the opportunity to do so is intense. In fact, presenting a paper at the more prestigious conferences is preferred to publication in a leading journal. Conference proceedings are not difficult to obtain if acquired in a timely fashion but can be hard to track down after more time elapses, particularly since many conference papers are never published in the journal literature.

Major Indexing and Abstracting Services

Indexes in computer science are used less heavily than in the other science disciplines. One researcher commented that many computer scientists mainly read their colleagues' preprints, often communicated by electronic mail, and ignore the rest of the world. They prefer to rely on word of mouth, in part because this seems to get them the most recent information and in part because systematic searching is time consuming. As experts in information management, some also have little faith in indexes and point out that access to indexed information is limited by the knowledge of those who categorize the data.

Print Indexes

ACM's *Computing Reviews*, a monthly publication begun in 1960, reviews current monographs and journal articles; it provides access by author, key word, subject, reviewer, and program name.

Computer and Control Abstracts, a section of *Science Abstracts*, published monthly by the Institution of Electrical Engineers, offers broad coverage with abstracts of current literature.

Science Citation Index (see **Biology** and **Physics**) provides broad access to computer science and related literature. Some, however, question its value for specialized research in computer science.

Electronic Indexes

By all accounts, electronic indexes are used as infrequently as print indexes. Few computer scientists wish to take the time to learn to do their own searching, and many find electronic indexes too expensive. However, some computer science librarians predict that if access were provided inexpensively on a campus network, this pattern would change.

Computing Reviews, along with the annual publication *ACM Guide to Computing Literature*, became available through MathSci in 1988 (see

Mathematics). *Computer and Control Abstracts* is accessible online as part of INSPEC (see **Physics**). Major indexes for other disciplines that also cover computer science material include NTIS and Compendex (see **Engineering**).

Current Awareness

Communication with colleagues is the principal means of staying up-to-date with current research. Many computer scientists use the library only infrequently to locate known items or articles listed in table-of-contents services. Conferences provide an important forum for exchange of information, and many computer scientists attend at least several each year. Electronic mail is almost universal among this group and the means for daily communication. Network newsletters and bulletin boards, which proliferate, are also popular.

Other Electronic Resources

Computer scientists do not need nonbibliographic databases such as those found in chemistry and biology. Instead of large databases, computer scientists need access to computing tools. Foremost among these are national and international computer networks linking them with researchers in this country and abroad. Networks allow for collaboration between computer scientists at different sites through electronic mail and the exchange of technical reports and programs via file transfer protocols. Networks are also beginning to provide access to servers that perform special services, such as the fabrication of VLSI chips.

As network capacity increases, networks will be used for even more demanding functions, including transmitting complex graphics and images, furnishing new computer architectures for testing, and processing and exchanging complex documents.

Standardization is an important issue in the electronic transmission of textual information. ACM and SIAM are developing systems for the electronic submission, review, editing, and production of articles for their journals as well as for retrospective conversion of remaining materials into electronic format. Further development of standards for machine-readable text will make it possible for one submission to be used to produce both a print publication and an electronic version capable of being searched.

Future Needs and Directions

Computer Science Literature

Journal literature is less important in computer science than in other disciplines. Computer scientists prefer electronic communication to print, but universities generally do not recognize electronic publications in making tenure and promotion decisions. Unless standards for tenure change to allow for electronic publication, the pressure on print journals to accommodate the increasing output of research may grow. At the same time, experimentation with electronic journals is likely to become more common.

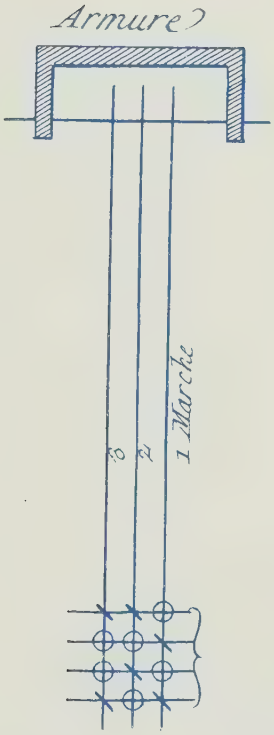
Technical reports will continue to be an important means of communicating research results. Access—preferably on a campus network—to the Stanford technical reports file would help to make technical reports

more widely available within the discipline. Faster and more comprehensive indexing of conference proceedings would also be useful.

Better and more rapid access to translations of work in Japanese, Russian, and Chinese is needed.

Electronic Resources

Projects to standardize information transfer protocols and to develop systems for the electronic transmission of articles should be supported. Increased network speeds are also necessary for opening up new opportunities for communication within computer science and with researchers in related fields.



Conclusion

Over the past 20 years the sciences have experienced an unprecedented rate of change. Change is evident across the science disciplines, in techniques and capabilities, in the structure of science disciplines and their interactions with other disciplines, and in the way information is created and disseminated. To a great extent, this ferment has been fueled by new technologies. Increasingly sophisticated instruments make possible experiments once impractical if not impossible. New information technologies, combining ever more powerful computers with ever wider and faster communication networks, generate and transfer huge amounts of data.

As new technologies continue to proliferate, there is every likelihood that the rate of change will continue unabated in the coming decades. The consequent financial, organizational, and intellectual challenges will be considerable. For those concerned with ensuring access to the information required for research, the most difficult task will be to keep pace with the staggering growth in research information of all types. This will be especially important in three areas.

Journal literature. As the output of scientific research has increased, so has the number of journals, many of them highly specialized. Thus, it is not surprising that many scientists point to the avalanche of journal literature, and the difficulty of keeping up with it, as the source of their greatest frustration. The burgeoning journal literature also presents difficulties for libraries. Those that once claimed to collect comprehensively can no longer do so; increasingly, they must look beyond their own collections to meet scientists' information requirements. Added to the problems posed by the higher volume of literature are alarmingly

higher costs, rising too rapidly for libraries to keep up. Further, libraries are more than ever in need of tools to help researchers sift through the mass of information available. As the body of information has grown, this task has become correspondingly more complex and more expensive.

Unpublished research. Just as the journal literature has grown, so has the volume of unpublished information such as preprints, technical reports, and conference papers, important to certain science disciplines and subdisciplines. These materials present problems of their own: knowledge of them tends to be confined to informal networks; they are difficult to identify, collect, and manage; and access is as haphazard as the management of them.

Primary data. Scientists must cope with scientific data that existed in a much smaller quantity or did not exist at all 20 years ago. Computers and increasingly sophisticated instruments such as linear accelerators and remote-sensing devices produce and store enormous quantities of data. Much of this information, although often produced at great expense and useful to a broad audience, is poorly managed and consequently underutilized.

Areas of Opportunity

For research libraries and other organizations with an interest in improving access to information for science research, some of the greatest opportunities lie in the following areas:

- Ensuring access to the primary literature through schemes such as cooperative lending arrangements and use of commercial or cooperative document delivery services.

- Facilitating communication of unpublished research results by working with the disciplines to provide more comprehensive bibliographic information than is now available; in some cases, provision of full-text information may also be warranted.

- Working with the disciplines, the government, and other data producers to provide better information about primary scientific data and distribute it on the national networks.

- Developing more integrated information environments that offer access to resources such as national and international computer networks for communication among researchers, bibliographic access to information on the latest published and unpublished research, information about primary data, and gateways to remote databases. More sophisticated systems might also offer the full text, with graphics, of the latest journal literature, software for manipulation of data, and collections of primary data.

- Preserving important older materials that have deteriorated.

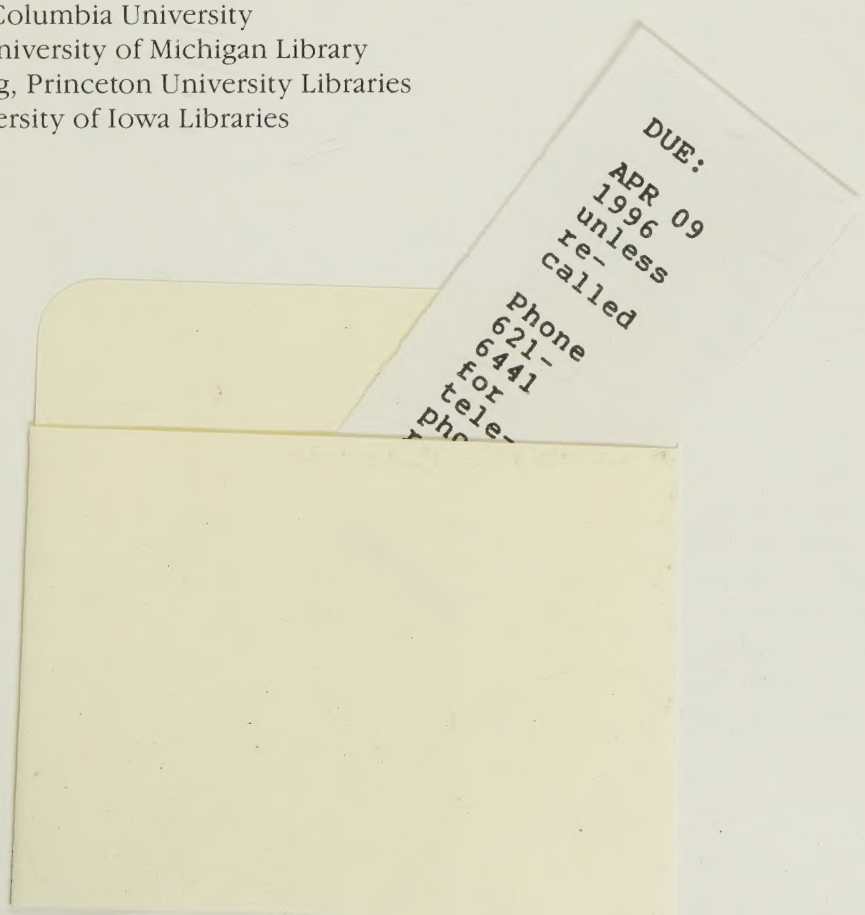
More than ever, it is clear that the free and unrestricted flow of information on which research depends is far beyond the capabilities of institutions or organizations working alone. In the future, communication in the sciences will be contingent on greater interdependence among all concerned—research libraries, the science disciplines, government agencies, commercial information providers, and scientists themselves. Cooperative approaches to the challenges presented by science information are possible—indeed, they are essential if scientists are to have access to the information vital to their research.

List of Persons Consulted in Surveying Needs of Science Disciplines

Louise Addis, Stanford Linear Accelerator
Faye Ajzenberg-Serlov, University of Pennsylvania
Robert Austin, Princeton University
Lisa Batzle, ARCO Oil and Gas Company
Grace Baysinger, Stanford University Libraries
Richard Beals, Yale University
Diana Beck, University of California at Berkeley Library
Stephan Beckwith, Cornell University
Sam Bedrosian, University of Pennsylvania
Charles Berry, University of Pennsylvania
Jerome Berson, Yale University
Linda Bickman, Genentech, Inc.
Richard Birnie, Dartmouth College
Arthur Bloom, Cornell University
Elaine Bouten, National Radio Astronomy Observatory
Katherine Branch, Yale University Library
James Burlitch, Cornell University
Blas Cabrera, Stanford University
Larry Carver, University of California at Santa Barbara
Robert Constable, Cornell University
Barbara Cooper, Cornell University
Peter Cziffra, Princeton University Libraries
Pierre Demarque, Yale University
Charlotte Derksen, Stanford University Libraries
Jackie Desoer, Chevron Research and Technology Company
Howard Dess, Princeton University Libraries
Frank Di Salvo, Cornell University
Carol Drum, University of Florida at Gainesville Libraries
Tom Duncan, University of California at Berkeley
Charles Early, Johns Hopkins University Libraries
L. Ehler, University of California at Davis
Rob Fesen, Dartmouth College
Gerald Fineberg, Columbia University
Steve Gass, Stanford University Libraries
Susan George, Dartmouth College Library
Larry Gerstein, University of California at Santa Barbara
Eleanor Goodchild, Stanford University Libraries
Joseph Goodman, Stanford University
Charles Grubmeyer, New York University
Joyce Hardy, Varian Associates
Isom Harrison, Clorox Technical Center
Martha Haynes, Cornell University
David Helfand, Columbia University
Herman Irving, Columbia University
Leo Hickey, Yale University
Wilson Ho, Cornell University
Helen Hoffman, Rutgers University Libraries

William Holser, Cornell University
Kimiyo Hom, University of California at Berkeley Library
Sarah How, Cornell University Library
Gary Johnson, Dartmouth College
H. Johnson, Cornell University
Alan Katritzky, University of Florida at Gainesville
Thomas Katz, Columbia University
J.G. Kaufman, Materials Property Data Network
Rafeal Kaufman, Yale University
William Kays, Stanford University
Jerry Kazden, University of Pennsylvania
Francis Kennedy, Dartmouth College
Claren Kidd, University of Oklahoma Libraries
Susan Klimley, Columbia University Libraries
Gillian Knapp, Princeton University
Donald Knuth, Stanford University
Walter Kohn, University of California at Santa Barbara
Dave Kristofferson, IntelliGenetics, Inc.
Rebecca Lasher, Stanford University Libraries
Alan Launer, Stanford University
Rita Lerner, American Institute of Physics
William Leveque, American Mathematical Society
Harry Llull, University of New Mexico Library
Henry Lowood, Stanford University Libraries
Daniel Lynch, Dartmouth College
John McCarthy, Stanford University
Gerald McGuire, Columbia University
Alan Meyers, University of Pennsylvania
Robert Michaelson, Northwestern University Library
Edward Morlok, University of Pennsylvania
George Mostow, Yale University
Steve Mount, Columbia University
Joan Murphy, Varian Associates
Jannette Naetzle, University of California at Davis
Jill Newby, Yale University Library
Mae Noffsinger, University of California at Davis
J.R. O'Neill, Chevron Chemical Company
Patricia O'Neill, Cornell University Library
Kim Parker, Yale University Library
Stanley Parsons, University of California at Santa Barbara
Stanton Peale, University of California at Santa Barbara
Sally Peters, Xerox Palo Alto Research Center
Leigh Phoenix, Cornell University
Henry Pinkham, Columbia University
Charles Prescott, Stanford Linear Accelerator Center
Alice Primack, University of Florida at Gainesville Libraries
John Rasmussen, University of California at Berkeley
Horst Richter, Dartmouth College
Bruce Rickborn, University of California at Santa Barbara
Steve Rocky, Cornell University Library
Ron Rodrigues, Stanford University Libraries

Frederick Rodriguez, Cornell University
James Rovnyak, University of Virginia
Halsey Royden, Stanford University
Danny Rye, Yale University
Gerald Salton, Cornell University
John Saylor, Cornell University Library
Fred Schneider, Cornell University
Mischa Schwartz, Columbia University
George Seligman, Yale University
Debonney Shoaf, University of California at Davis
Wendy Silk, University of California at Davis
David Simpson, Columbia University
William Sofer, Rutgers University
Ronald Standler, Pennsylvania State University
Cynthia Steinke, University of Minnesota Libraries
Dan Stinebring, Princeton University
Barbara Stitt, New York University
John Terborgh, Princeton University
Donald Turcott, Cornell University
Betty Tyson, Westinghouse Electric Corporation
David Tytler, Columbia University
Steve Vavasis, Cornell University
Matthew Vernon, Columbia University
Rosalind Walcott, State University of New York at Stony Brook Library
Arthur Walker, Stanford University
Camille Wanat, University of California at Berkeley Library
Gary Wegner, Dartmouth College Library
Jack Weigel, University of Michigan Library
Beth Weil, University of California at Berkeley Library
Yechiam Yemini, Columbia University
Patricia Yokum, University of Michigan Library
Helen Zimmerberg, Princeton University Libraries
Louise Zipp, University of Iowa Libraries



3 9001 02943 5206

Note on Illustrations

All illustrations are taken from *Encyclopédie, ou Dictionnaire raisonné des sciences, des arts et des métiers, par une société de gens de lettres. Mis en ordre et publié par M. Diderot . . . et quant à la partie mathématique, par M. d'Alembert . . .* Paris, Briasson [etc.], 1751–65. They are reproduced through the kind assistance of the Department of Special Collections and Archives, Stanford University Libraries.

Diderot's famous *Encyclopédie* was a monumental effort to classify and describe the scientific and technological knowledge of the time. The 28-volume work (17 volumes of text and 11 volumes of illustrations) surpassed all previous encyclopedias and dictionaries and is considered a landmark of scientific scholarship to this day.

Cover: "Essai d'une distribution généalogique des sciences et des arts principaux." **Inside:** Introduction and Conclusion—"Le laboratoire chimique"; Physics—"Machine inventée pour mesurer l'électricité"; Chemistry—"Appareil pour la distillation de l'huile de vitriol"; Biology—"Actinia-sociata, ou animal-fleur"; Astronomy—"Hémisphere boréal"; Geosciences—"Autre sommet du Vésuve pendant une petite éruption"; Engineering—"Mouton à cheval sur bateaux"; Mathematics—"Relative aux mots arcs semblables, hauteur, angle, etc. . . ."; Computer Science—"Soierie - cantre à la Lionoise vue en perspective."

The Research Libraries Group, Inc.